1971

Screening Electrocardiograms Using Hotelling's Method of Principle Components

Edward G. Schwartz

SCREENING ELECTROCARDIOGRAMS USING HOTELLING'S METHOD OF PRINCIPLE COMPONENTS

BY

EDWARD G. SCHWARTZ

A thesis submitted
in partial fulfillment of the requirements for the
degree Master of Science, Major in Electrical
Engineering, South Dakota
State University

1971

SCREENING ELECTROCARDIOGRAMS USING HOTELLING'S

METHOD OF PRINCIPLE COMPONENTS

This thesis is approved as a creditable and independent investigation by a candidate for the degree, Master of Science, and is acceptable as meeting the thesis requirements for this degree, but without implying that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Thesis Advisor Date

Head, Electrical Engineering Department Date

ACKNOWLEDGEMENTS

The author wishes to express his appreciation and gratitude to Dr. Kurtenbach for his guidance and advice throughout the author's graduate work and in the research and preparation of this thesis.

EGS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF GRAPHS

LIST OF FIGURES

INTRODUCTION

Considerable use is presently being made of computers to classify electrocardiograms (ECGs) into various heart defect categories. An example of such a program is the "Electrocardiographic Analysis Program"[8] developed jointly by the Mayo Clinic and IBM. In one of Dr. Caceres published reports[4], he tells of using a computer program which involves 5,000 instructions. Dr. Caceres does not say whether this is a commercial program or one of his own. His description does place this program in the same general category as IBM's. These programs have the common criteria of examining the ECG for the same information that a cardiologist uses to analyze an ECG. This necessitates using many instructions and large memory storage facilities; thus, a large computer is necessary.

It would be most desirable to have a computer method of analyzing ECGs that would utilize a small special purpose computer and eliminate the use of the large expensive computer. A second choice would be a small special purpose computer capable of sorting the cardiograms into the categories of normal and possibly abnormal. The possibly abnormal ECGs would then be analyzed on the large computer for diagnostic purposes.

Bailey[3], in his thesis, examined the possibility of changing the ECG from a time-varying plot to a statistical plot not involving time. This probability density function was analyzed to determine the ECG

category. He was able to show that correlation exists between the probability density function and the ECG classification of normal versus possibly abnormal.

Young and Huggins[20] tried approximating the ECG as a twelve dimensional vector utilizing the orthonormal exponentials as the basis vectors. The coefficients of these basis vectors were used to categorize the ECGs. Seventy-five percent correct diagnosis was obtained on 65 cardiograms; 53 of these cardiograms comprised the training set. It is of interest to repeat what these authors call a fundamental assumption justifying this type of approach.

> "Since the physicians are able to distinguish different pathological categories from the similarities and the dissimilarities of ECG waveforms, and since the waveform corresponds to direction in signal space, it is logical to conclude that the signal space may be separated into several subspaces, each subspace corresponding to a pathological category. A transformation may be found to relate the subspaces with the pathological categories. The ECG signal vector which falls into a certain subspace may then be considered as belonging to that category."

This thesis utilizes the above fundamental assumption and attempts to assign ECGs to one of two classes (normal or possibly abnormal) on the basis of examples given for each class; thus, a pattern recognition

problem[5] is the result. To generate the basis vectors that span the signal space, the Hotelling Method[10] is used. This method utilizes a linear transformation to obtain orthogonal basis vectors (the uncorrelated random variables of statistics). The vectors are generated so that the first vector yields the largest possible variance; the second vector yields the next largest possible variance; etc. Variance is a direct measure of the amount of information contained in the vector[11].

Mattson and Damman[12] were able to segregate voice patterns by using only the first vector obtained by the Hotelling Method. The theory used by these gentlemen to acquire the vector was slightly different from Hotelling's approach; their theory produced only the vector yielding the largest variance. A better classification might have resulted if more than one vector had been considered.

THE PATTERN RECOGNITION PROBLEM

Pattern recognition is the development of equipment and techniques to implement the automatic recognition of patterns[16]. This definition omits categories requiring human attention such as photo interpretation.

The block diagram of a general pattern recognition device (pattern classifier) is as follows:

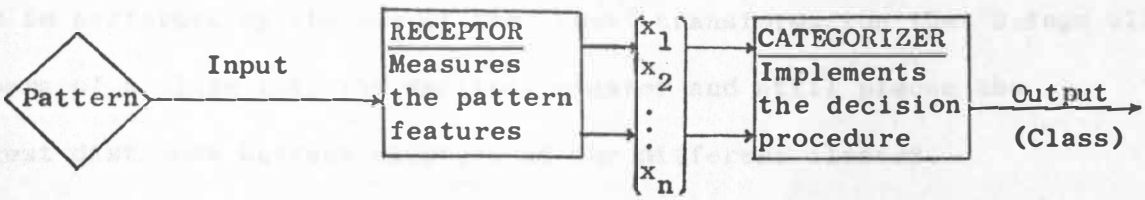


Figure 1. Block Diagram of a Pattern Classifier

The feature vector is the n different values utilized by the categorizer to make a decision (hopefully the correct one). These n different values are obtained by making measurements of n different properties on the pattern, and assigning a real value, x_i, to each property.

The performance of a pattern classifier is closely related to the measurements taken by the classifier[9]. Usually there exists the possibility of taking many more measurements on the pattern than the handling capabilities of the pattern classifier will allow. The "feature selection problem" (sometimes referred to as preprocessing, filtering or prefiltering, feature or measurement extraction, or dimensionality

reduction[13]) is the optimum choice of quantities to be measured and manipulated to create the feature vector. This manipulation can be considered as the transformation from the original measurement or observation space to the feature space.

Each of the n components of the feature vector represents some property of the pattern. Each pattern can therefore be represented by a vector (or point) in n-dimensional space. Working on the premise that similar patterns will have similar properties, one expects the vectors representing one class of patterns to cluster in feature space. This is performed by the use of the "best" transformation that brings all members of a class into the smallest cluster and still places the largest distances between clusters of the different classes.

A set of previously classified patterns is applied to the receptor with its output being the feature vectors. These vectors are examined, and the decision procedure of the categorizer is "optimized" to give the best results. The primary goal of a pattern classifier is to achieve a high recognition rate on new data[13]; this is quite often approximated as that procedure yielding the highest recognition rate on the training set.

This classification procedure is implemented with a <u>discriminant function</u>[14]. The discriminant function is a real valued function which combines the values from the feature vector so that a single number results. On the basis of this number, the classification is made. Discriminant functions may be sorted into two general categories: linear and nonlinear.

An example[14] of a linear discriminant function is

$$u(\underset{\sim}{x}) = \underset{\sim}{V}' \underset{\sim}{X} + v_o \qquad (1)$$

This may be utilized in prediction as follows: choose class C_1 if $u(\underset{\sim}{x}) \geq \Theta$ otherwise choose class C_2. Utilizing the minimum distance to the means criterion would be letting $u(\underset{\sim}{x}) = |\underset{\sim}{x}-\underset{\sim}{\bar{x}}_2|$ and $\Theta = |\underset{\sim}{x}-\underset{\sim}{\bar{x}}_1|$. $\underset{\sim}{\bar{x}}_1$ and $\underset{\sim}{\bar{x}}_2$ are the mean matrices. A similar criterion is the minimum distance to the nearest neighbor. The correlation criterion is where $\underset{\sim}{V} = \underset{\sim}{\bar{x}}_1 - \underset{\sim}{\bar{x}}_2$ and $\Theta = \frac{1}{2}(\underset{\sim}{x}_1-\underset{\sim}{x}_2)'(\underset{\sim}{x}_1+\underset{\sim}{x}_2)$. Maximum likelihood criterion[13] utilizing Bayes' formula for conditional probabilities results in a linear function when the x's are statistically independent, when the x's are binary, and when Gaussian distributions have identical covariance matrices. The minimax decision rule (Anderson-Bahadur formula)[13] yields a linear function when applied to Gaussian Distributions with unequal covariance matrices. Discriminant analysis may be used where the form of the probability density function is not known. A class of interesting linear discriminant functions are the potential functions[13]. These are of the form

$$\Psi(\underset{\sim}{x}) = \Sigma W_i \phi_i (\underset{\sim}{x}) \qquad (2)$$

The function is positive for patterns in the first class and negative for those in the second class.

In the nonlinear discriminant function category is the regular Gaussian distribution where the optimum separation surface (or hyperplane if the n dimensions exceeds 2) is defined by letting the probability density function of the first class equal the probability density function of the second class. Again this is a type of nearest neighbor criterion.

Both the linear and nonlinear categories can utilize trainable discriminant functions where the weight vector is changed incrementally as more and more patterns are applied to the classifier.

Another type is the sequential decision categorizer. By this method the next point inspected and weighted depends on the results obtained from inspecting previous points. The decision is based on the weighted values obtained from the various points.

In formulating the solution to the pattern recognition problem various assumptions have been made[17]. The first assumption is that the n-dimensional space chosen to display the feature vector is a sufficiently complete model to contain enough information on the common properties of the various classes to facilitate classification. Generally, one must rely on engineering judgement and intuition to determine whether the model is sufficiently complete. Usually, this decision can be made with considerable confidence.

A second assumption is that the transformation of the data from observation space to feature space is by the "best" method. Transformations can be obtained for some certain sets of criteria that define the "best" method. Generally, one has to use engineering judgement to settle for a "good" transformation.

Another assumption is that the "best" discriminant function can be specified. Again, good engineering judgement enters the picture.

A fourth assumption is that the training set of patterns is large enough to be classified as a representative set. This, also, must usually be satisfied by engineering judgement.

HOTELLING METHOD[1,10]

Consider a random vector

$$\underset{\sim}{X} = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \vdots \\ x_p \end{pmatrix} \tag{3}$$

attached to each member of a population. The components, $x_1,\ldots,x_p$, are p variables or measurements taken on each population member. These measurements are in general correlated. Assume that the mean vector of $\underset{\sim}{X}$ is $\underset{\sim}{0}$ and the covariance matrix

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \tag{4}$$

exists. By definition, the covariance is given by

$$\sigma_{ij} = \frac{1}{N} \sum_{l=1}^{N} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j) \tag{5}$$

In the present case the means are zero, therefore

$$\sigma_{ij} = \frac{1}{N} \sum_{l=1}^{N} x_{il} x_{jl} \tag{6}$$

Thus, the covariance matrix is the expectation values of the various $x_i x_j$ products.

A set of orthonormal, linear combinations of the x's is desired. This will remove the correlation between the variables and give a unique description of the population member.

Let the desired linear combination be written as $\underset{\sim}{B}'\underset{\sim}{X}$ where

$$B = \begin{pmatrix} B_1 \\ . \\ . \\ . \\ B_p \end{pmatrix} \qquad (7)$$

and

$$\underset{\sim}{B}'\underset{\sim}{B} = 1 \qquad (8)$$

The prime indicates the transpose.

One finds the variance of $\underset{\sim}{B}'\underset{\sim}{X}$ by

$$\text{Variance of } \underset{\sim}{B}'\underset{\sim}{X} = E(\underset{\sim}{B}'\underset{\sim}{X})^2 = E\ \underset{\sim}{B}'\underset{\sim}{X}\underset{\sim}{X}'\underset{\sim}{B} = \underset{\sim}{B}'\underset{\sim}{\Sigma}\underset{\sim}{B} \qquad (9)$$

The problem now is to find $\underset{\sim}{B}'$ such that $\underset{\sim}{B}'\underset{\sim}{X}$ has maximum variance subject to the constraint of (8). This can be performed by the Lagrange multiplier technique. Form

$$\phi_1 = \underset{\sim}{B}'\underset{\sim}{\Sigma}\underset{\sim}{B} - \lambda(\underset{\sim}{B}'\underset{\sim}{B} - 1) \qquad (10)$$

Now take the vector of partial derivatives, $\frac{\partial \phi}{\partial \underset{\sim}{B}}$, and set it equal to $\underset{\sim}{0}$ to acquire a $\underset{\sim}{B}$ to maximize $\underset{\sim}{B}'\underset{\sim}{\Sigma}\underset{\sim}{B}$. So

$$\frac{\partial \phi_1}{\partial \underset{\sim}{B}} = 2\underset{\sim}{\Sigma}\underset{\sim}{B} - 2\lambda\underset{\sim}{B} = \underset{\sim}{0} \qquad (11)$$

or

$$(\underset{\sim}{\Sigma} - \lambda \underset{\sim}{I})\ \underset{\sim}{B} = \underset{\sim}{0} \qquad (12)$$

For $\underset{\sim}{B}$ to have a nontrivial solution $(\underset{\sim}{\Sigma} - \lambda\underset{\sim}{I})$ must not have an inverse. This means that

$$|\underset{\sim}{\Sigma} - \lambda\underset{\sim}{I}| = 0 \qquad (13)$$

This polynomial has p roots: $\lambda_1, \lambda_2, \ldots, \lambda_p$ (the λ's may be called characteristic values, eigenvalues, proper values, or latent roots)[6,7].

Substituting (13) into (9) one obtains the fact that the variance of $\underset{\sim}{B}'\underset{\sim}{X}$ is equal to λ. For maximum variance one utilizes the largest root (call it λ_1).

Let $\underset{\sim}{B}_1$ be a normalized solution of (12). So

$$\underset{\sim}{U}_1 = \underset{\sim}{B}_1'\underset{\sim}{X} \tag{14}$$

is a normalized linear combination of maximum variance.

Next a new $\underset{\sim}{B}$ is desired that will give $\underset{\sim}{B}'\underset{\sim}{X}$ a maximum variance and is orthogonal in statistical and geometric sense with $\underset{\sim}{U}_1$. Thus,

$$0 = E\,\underset{\sim}{B}'\underset{\sim}{X}\underset{\sim}{U}_1' = E\,\underset{\sim}{B}'\underset{\sim}{X}\underset{\sim}{X}'\underset{\sim}{B}_1 = \underset{\sim}{B}'\underset{\sim}{\Sigma}\,\underset{\sim}{B}_1 = \lambda_1\underset{\sim}{B}'\underset{\sim}{B}_1 \tag{15}$$

In other words, it is necessary to have

$$\underset{\sim}{B}'\underset{\sim}{B}_1 = 0 \tag{16}$$

Again, the Lagrange multiplier technique can be employed;

$$\phi_2 = \underset{\sim}{B}'\underset{\sim}{\Sigma}\,\underset{\sim}{B} - \lambda(\underset{\sim}{B}'\underset{\sim}{B} - 1) - 2\,\gamma_1\underset{\sim}{B}'\underset{\sim}{\Sigma}\,\underset{\sim}{B}_1 \tag{17}$$

The vector of partial derivatives is set equal to zero.

$$\frac{\partial \phi_2}{\partial \underset{\sim}{B}} = 2\underset{\sim}{\Sigma}\,\underset{\sim}{B} - 2\lambda\underset{\sim}{B} - 2\gamma_1\underset{\sim}{\Sigma}\,\underset{\sim}{B}_1 = \underset{\sim}{0} \tag{18}$$

Premultiply by $\underset{\sim}{B}_1'$.

$$\underset{\sim}{B}_1'\underset{\sim}{\Sigma}\,\underset{\sim}{B} - \lambda\underset{\sim}{B}_1'\underset{\sim}{B} - \gamma_1\underset{\sim}{B}_1'\underset{\sim}{\Sigma}\underset{\sim}{B}_1 = \underset{\sim}{0} \tag{19}$$

or

$$\gamma_1\,\lambda_1 = 0 \tag{20}$$

Thus

$$\gamma_1 = 0 \tag{21}$$

One now solves for the new $\underset{\sim}{B}$ from

$$(\underset{\sim}{\Sigma} - \lambda_2\underset{\sim}{I})\underset{\sim}{B}_2 = 0 \tag{22}$$

and forms the new linear combination

$$\underset{\sim}{U}_2 = \underset{\sim}{B}_2'\underset{\sim}{X} \tag{23}$$

This is continued until one has the full set of p vectors of $\underset{\sim}{U}_i$. These vectors are called the principal components (sometimes referred to as characteristic vectors or irreducible representations).

Now form a new pxp matrix $\underset{\sim}{B}$ composed of the various $\underset{\sim}{B}_i$ column matrices.

$$\underset{\sim}{B} = (\underset{\sim}{B}_1 \ldots \underset{\sim}{B}_p) \tag{24}$$

The matrix $\underset{\sim}{\Lambda}$ may also be formed:

$$\underset{\sim}{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \ldots & \lambda_p \end{pmatrix} \tag{25}$$

From the fact that

$$\underset{\sim}{\Sigma}\underset{\sim}{B}_r = \lambda_r \underset{\sim}{B}_r \tag{26}$$

one may write

$$\underset{\sim}{\Sigma}\underset{\sim}{B} = \underset{\sim}{B}\underset{\sim}{\Lambda} \tag{27}$$

Since

$$\underset{\sim}{B}_r\underset{\sim}{B}_s' = \delta_{rs} \tag{28}$$

one has

$$\underset{\sim}{B}'\underset{\sim}{\Sigma}\underset{\sim}{B} = \underset{\sim}{\Lambda} \tag{29}$$

If

$$\underset{\sim}{U} = \underset{\sim}{B}'\underset{\sim}{X} \tag{30}$$

then by (9) and (13) the covariance matrix of $\underset{\sim}{U}$ is $\underset{\sim}{\Lambda}$, and $\underset{\sim}{U}$ is the vector of principal components of $\underset{\sim}{X}$. A computer oriented procedure to find $\underset{\sim}{U}$ and $\underset{\sim}{\Lambda}$ from $\underset{\sim}{\Sigma}$ is given in appendix A.

DEVELOPMENT OF THE CLASSIFICATION PROCEDURE

The data for this thesis has been obtained from the University of Nebraska College of Medicine. This data, which has been placed on computer tape, is from the Frank lead type of cardiogram. Each consecutive set of X, Y, and Z data points was separated in time by 0.004 seconds. Each data value could be converted to volts by multiplying by +0.00015259.

A reference point had to be chosen which was reproducible in each cardiogram. The point of greatest negative slope on an ECG signal occurs after the R-wave peak and before the S-wave peak. This value never varies over 30% in any one subject[2,18]. Steinberg[18] points out that the heartbeat interval is always greater than 0.64 seconds. The cardiogram was divided into 0.64 second intervals and the point of greatest slope in each interval was located. The first four values that were found to be within 30% of the most negative value were assumed to correspond to similar points in each heartbeat. If the pulse durations of the three heartbeats varied by less than 10% the values of corresponding data points were averaged together; thus, an average ECG was formed. If the pulse duration variations exceeded 10% then the data from that heartbeat with a pulse rate closest to 72 beats per minute was used for analysis.

The X-lead is reported to possess the most information[3]. X-lead data located 0.28 seconds before to 0.32 seconds after the point of

maximum slope was selected to be utilized. This time portion of the ECG contains the P, Q, R, S, and T waves. The pulse rate had to be between 62 pulses per minute and 82 per minute for the ECG to be used in the analysis. A flow-diagram of the subprogram used to acquire the data to be analyzed is presented in appendix B.

The values of the data points located .008 seconds apart were averaged for the normal cardiograms, for the abnormal cardiograms, and for the two together. The results of this averaging are recorded in Table 1; position 35 is the location of maximum negative slope. A total of 32 normal and 28 abnormal or marginal cardiograms were used to acquire the averages of Table 1.

The averages to be used in Hotelling's method calls for the averages to be those of the actual population under surveillance. The data has almost a fifty-fifty distribution between normal and abnormal cardiograms. It was decided to use averages based on 25% abnormal and 75% normal; this distribution was assumed to more closely resemble the distribution encountered in a doctor's office. The averages based on this distribution are listed in Table 2.

It was decided to place a zero line through each cardiogram. This was done by averaging the values of each of the 75 points and then subtracting this value from each point. The effect of this procedure was to move the cardiogram's baseline such that the area under the curve above the baseline is equal to the area under the curve below the baseline.

TABLE 1

AVERAGE CARDIOGRAM DATA VALUES

position	normal	abnormal	no distinction
1	-1334	-1684	-1391
2	-1179	-1724	-1433
3	-1161	-1709	-1417
4	-1160	-1725	-1424
5	-1138	-1767	-1431
6	-1165	-1751	-1438
7	-1146	-1744	-1425
8	-1154	-1730	-1423
9	-1118	-1719	-1399
10	-1097	-1723	-1389
11	-1047	-1707	-1355
12	-1024	-1708	-1343
13	-1027	-1630	-1308
14	-1062	-1515	-1273
15	-1011	-1460	-1221
16	- 981	-1614	-1276
17	- 948	-1691	-1295
18	- 945	-1780	-1335
19	- 990	-1797	-1367
20	- 966	-1883	-1394
21	- 996	-1995	-1462

TABLE 1 (cont.)

position	normal	abnormal	no distinction
22	- 984	-2081	-1496
23	-1045	-2095	-1535
24	-1059	-2095	-1543
25	-1096	-2049	-1541
26	-1057	-1925	-1462
27	-1084	-1901	-1465
28	-1048	-1863	-1428
29	-1121	-1873	-1472
30	-1109	-1747	-1407
31	- 396	-1039	- 696
32	1461	468	998
33	3656	2461	3099
34	3897	2778	3375
35	772	35	428
36	-1237	-1858	-1527
37	-1566	-2451	-1979
38	-1386	-2469	-1891
39	-1197	-2365	-1742
40	-1128	-2296	-1673
41	-1208	-2188	-1665
42	-1247	-2188	-1686
43	-1219	-2085	-1624

TABLE 1 (cont.)

position	normal	abnormal	no distinction
44	-1218	-1977	-1572
45	-1172	-1913	-1518
46	-1167	-1866	-1493
47	-1133	-1814	-1450
48	-1121	-1833	-1453
49	-1108	-1837	-1448
50	-1091	-1791	-1418
51	-1038	-1731	-1361
52	-1021	-1668	-1323
53	- 886	-1637	-1237
54	- 856	-1624	-1215
55	- 826	-1641	-1206
56	- 772	-1589	-1153
57	- 679	-1647	-1131
58	- 580	-1606	-1059
59	- 443	-1566	- 967
60	- 339	-1486	- 874
61	- 198	-1370	- 745
62	- 112	-1331	- 681
63	- 36	-1263	- 609
64	- 15	-1158	- 548
65	- 26	-1206	- 577

TABLE 1 (cont.)

position	normal	abnormal	no distinction
66	- 81	-1097	- 555
67	- 141	-1072	- 576
68	- 235	-1057	- 619
69	- 294	-1057	- 650
70	- 452	-1176	- 790
71	- 664	-1345	- 982
72	- 761	-1445	-1080
73	- 877	-1527	-1180
74	- 925	-1599	-1239
75	- 982	-1603	-1271

TABLE 2

MEAN VALUE OF POINTS REPRESENTING A 75-25 PERCENT DISTRIBUTION

POINT	VALUE	POINT	VALUE	POINT	VALUE
1	-1422	21	-1246	41	-1453
2	-1315	22	-1258	42	-1482
3	-1298	23	-1308	43	-1436
4	-1301	24	-1318	44	-1408
5	-1295	25	-1334	45	-1357
6	-1312	26	-1274	46	-1342
7	-1296	27	-1288	47	-1303
8	-1298	28	-1252	48	-1299
9	-1268	29	-1309	49	-1290
10	-1254	30	-1269	50	-1266
11	-1212	31	- 557	51	-1211
12	-1195	32	1213	52	-1183
13	-1178	33	3357	53	-1074
14	-1175	34	3617	54	-1048
15	-1123	35	588	55	-1030
16	-1139	36	-1392	56	- 976
17	-1134	37	-1787	57	- 921
18	-1154	38	-1657	58	- 837
19	-1192	39	-1489	59	- 724
20	-1195	40	-1420	60	- 626

TABLE 2 (cont.)

POINT	VALUE	POINT	VALUE	POINT	VALUE
61	- 491	66	- 335	71	- 834
62	- 417	67	- 374	72	- 932
63	- 343	68	- 441	73	-1040
64	- 301	69	- 485	74	-1094
65	- 321	70	- 633	75	-1137

The sixty cardiograms originally used to compute averages were now used to determine the covariance matrix. This matrix was rotated by the use of the Jacobi method (see appendix) to yield a diagonal matrix. The rotated matrix was the matrix of characteristic values; an associated matrix of characteristic vectors was also created.

The coefficients associated with each term of the first four characteristic vectors are listed in Table 3. The characteristic values (the variances) associated with these first four characteristic vectors were 26,543,408; 13,514,495; 6,316,394; and 3,706,386. The next characteristic value was less than 10% of the largest and therefore was not considered. Graph 1 is a plot of the characteristic values versus the characteristic vector number. Selection of these four vectors was based on the proportionality between a vector's information content and its variance.

Data of the 60 cardiograms (the training set) was impressed on the four characteristic vectors. An average vector length was found for each vector under the normal category and under the abnormal category. This allowed a point to be found in four-space that was the average of the normal ECGs, and another point was found that was the average of the abnormal ECGs. Coordinates of these points were (273.9, -106.6, -19.3, 84.5) for the normal ECGs and (187.9, -237.2, 71.9, -87.3) for the abnormals.

Each cardiogram was now impressed on the characteristic vectors; thus, a point in four-space was found as the location of the cardiogram.

TABLE 3

CHARACTERISTIC VECTOR COEFFICIENTS OBTAINED FROM

75% NORMAL-25% ABNORMAL COVARIANCE MATRIX

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
1	.188	-.113	.029	-.054
2	.159	-.095	.015	-.046
3	.138	-.087	.016	-.051
4	.122	-.078	-.003	-.033
5	.135	-.084	-.007	-.033
6	.125	-.088	.009	-.045
7	.137	-.087	-.001	-.029
8	.132	-.077	-.002	-.019
9	.124	-.078	-.003	-.018
10	.120	-.074	-.003	-.030
11	.124	-.070	-.027	.000
12	.129	-.070	-.033	.007
13	.124	-.086	.006	-.026
14	.105	-.115	.070	-.097
15	.101	-.132	.106	-.105
16	.114	-.088	.009	-.042
17	.129	-.038	-.084	.029
18	.142	.010	-.161	.087
19	.140	.017	-.190	.122

TABLE 3 (cont.)

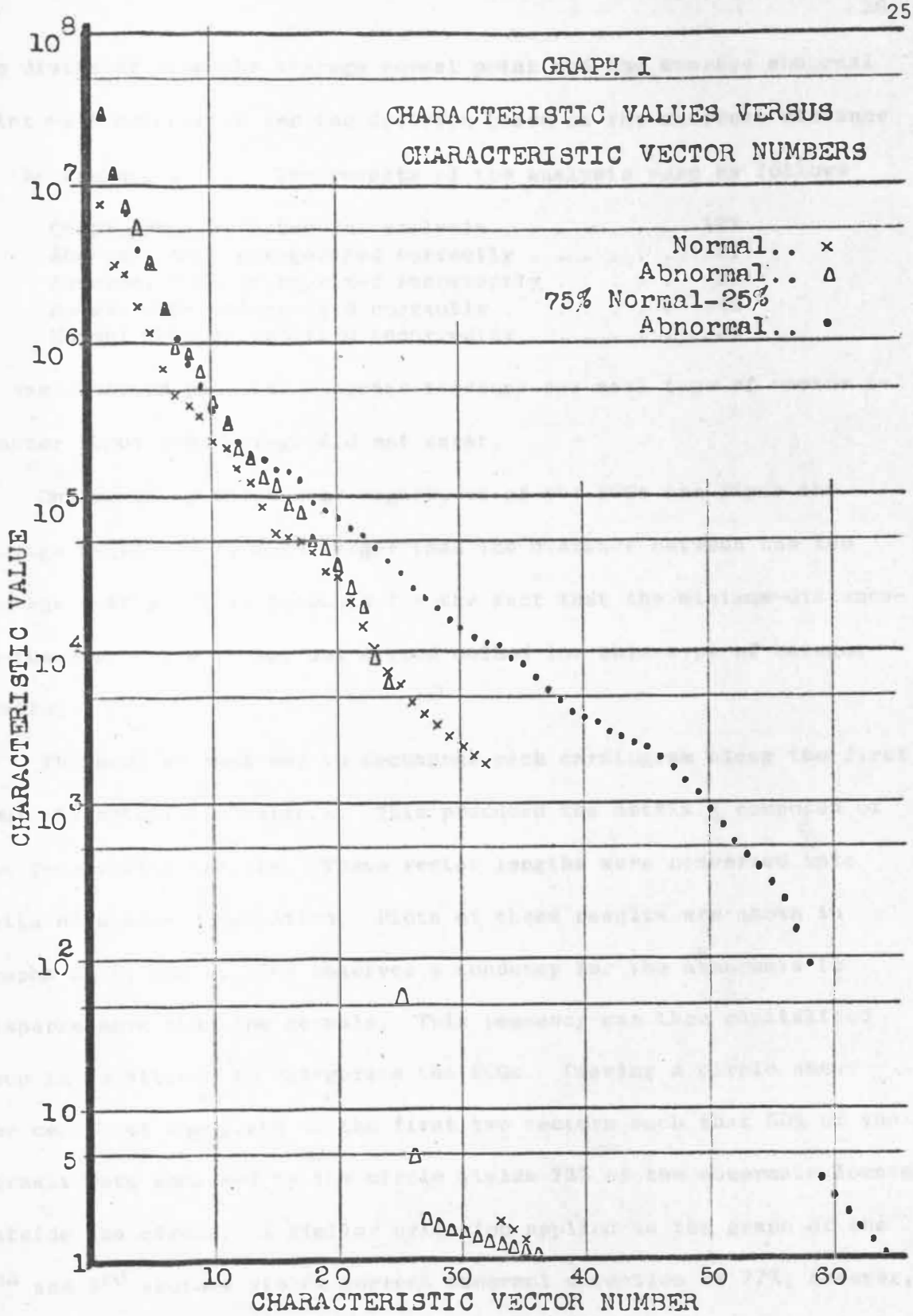
Coeff. No.	Vector Number. 1st	2nd	3rd	4th
20	.137	.038	-.215	.128
21	.138	.032	-.222	.142
22	.129	.027	-.196	.126
23	.122	.012	-.183	.095
24	.117	.004	-.147	.074
25	.104	-.022	-.089	.031
26	.084	-.043	-.042	-.008
27	.070	-.052	.003	-.020
28	.072	-.057	.011	-.037
29	.071	-.052	.011	-.069
30	.086	-.033	.080	-.205
31	.095	.094	.132	-.375
32	.105	.343	.129	-.449
33	.124	.559	.100	-.179
34	.120	.523	.073	.210
35	.076	.268	.202	.340
36	.026	-.006	.249	.328
37	-.008	-.037	.250	.227
38	-.030	-.035	.250	.168
39	-.044	-.033	.220	.139
40	-.047	-.022	.189	.137

TABLE 3 (cont.)

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
41	-.048	-.028	.191	.095
42	-.060	-.028	.151	.104
43	-.071	-.034	.138	.041
44	-.076	-.040	.105	-.001
45	-.085	-.044	.088	-.050
46	-.094	-.047	.082	-.048
47	-.100	-.054	.077	-.063
48	-.107	-.050	.065	-.043
49	-.108	-.053	.064	-.052
50	-.113	-.048	.074	-.065
51	-.113	-.050	.057	-.055
52	-.121	-.058	.049	-.078
53	-.127	-.053	.045	-.052
54	-.129	-.040	.042	-.042
55	-.120	-.029	.025	-.052
56	-.130	-.015	.025	-.050
57	-.117	.001	-.016	-.035
58	-.125	.014	-.035	-.029
59	-.131	.026	-.058	-.026
60	-.121	.042	-.089	-.009
61	-.133	.052	-.083	-.018
62	-.126	.061	-.116	.003

TABLE 3 (cont.)

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
63	-.131	.062	-.128	.007
64	-.130	.070	-.134	-.016
65	-.132	.063	-.127	.006
66	-.130	.065	-.140	-.010
67	-.132	.060	-.140	-.003
68	-.135	.039	-.150	.004
69	-.142	.025	-.144	.028
70	-.133	.028	-.135	.022
71	-.122	.006	-.102	.010
72	-.125	-.003	-.074	.000
73	-.120	.005	-.062	.017
74	-.121	-.002	-.065	.010
75	-.128	-.014	-.038	.021


GRAPH I
CHARACTERISTIC VALUES VERSUS
CHARACTERISTIC VECTOR NUMBERS
Normal.. ×
Abnormal.. Δ
75% Normal-25%
Abnormal.. •
CHARACTERISTIC VALUE
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10
5
1
10
20
30
40
50
60
CHARACTERISTIC VECTOR NUMBER

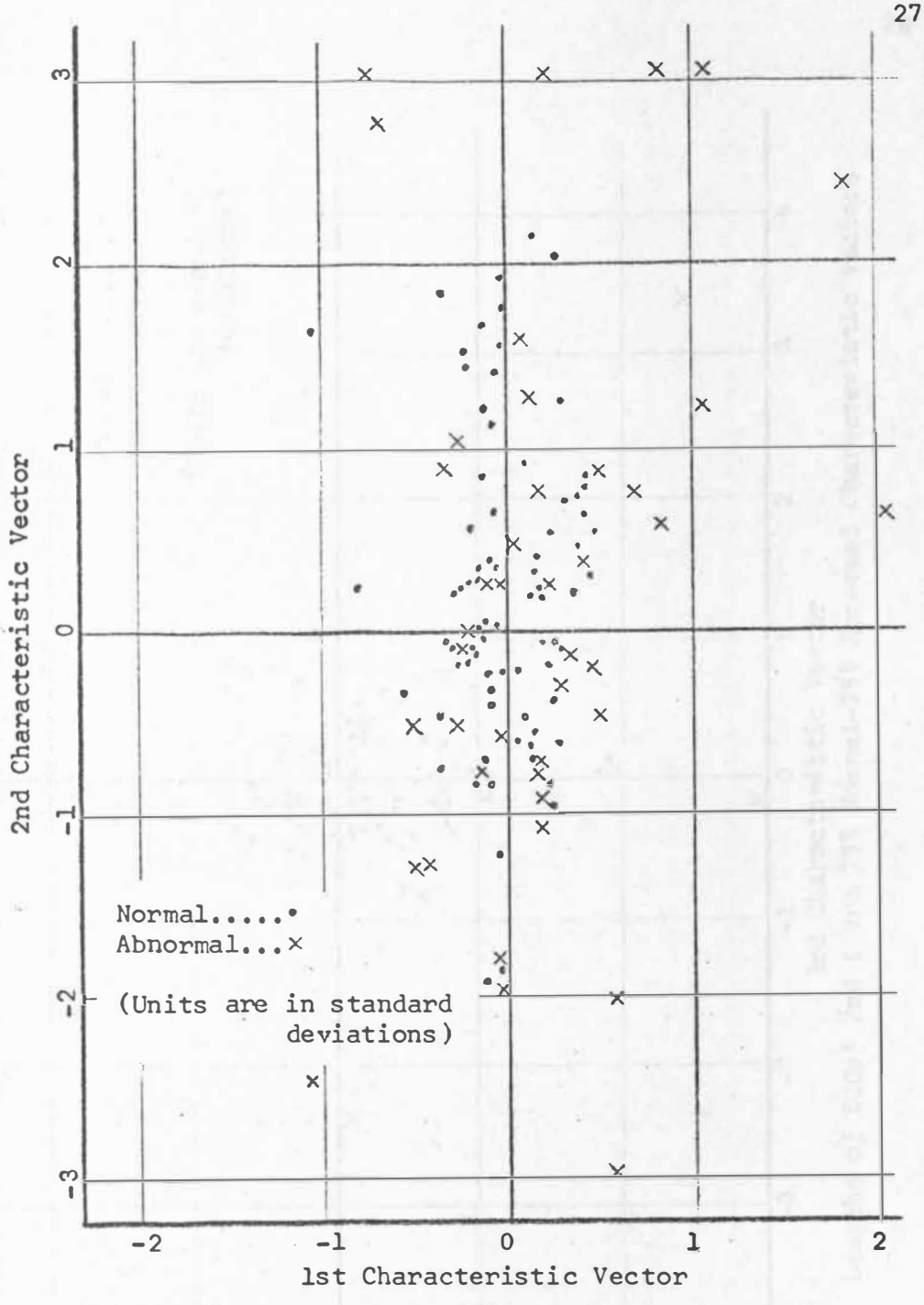
The distances from the average normal point and the average abnormal point were calculated and the decision based on the shortest distance to the average point. The results of the analysis were as follows:

Cardiograms selected for analysis	132
Abnormal ECGs categorized correctly	21
Abnormal ECGs categorized incorrectly	28
Normal ECGs categorized correctly	48
Normal ECGs categorized incorrectly	35

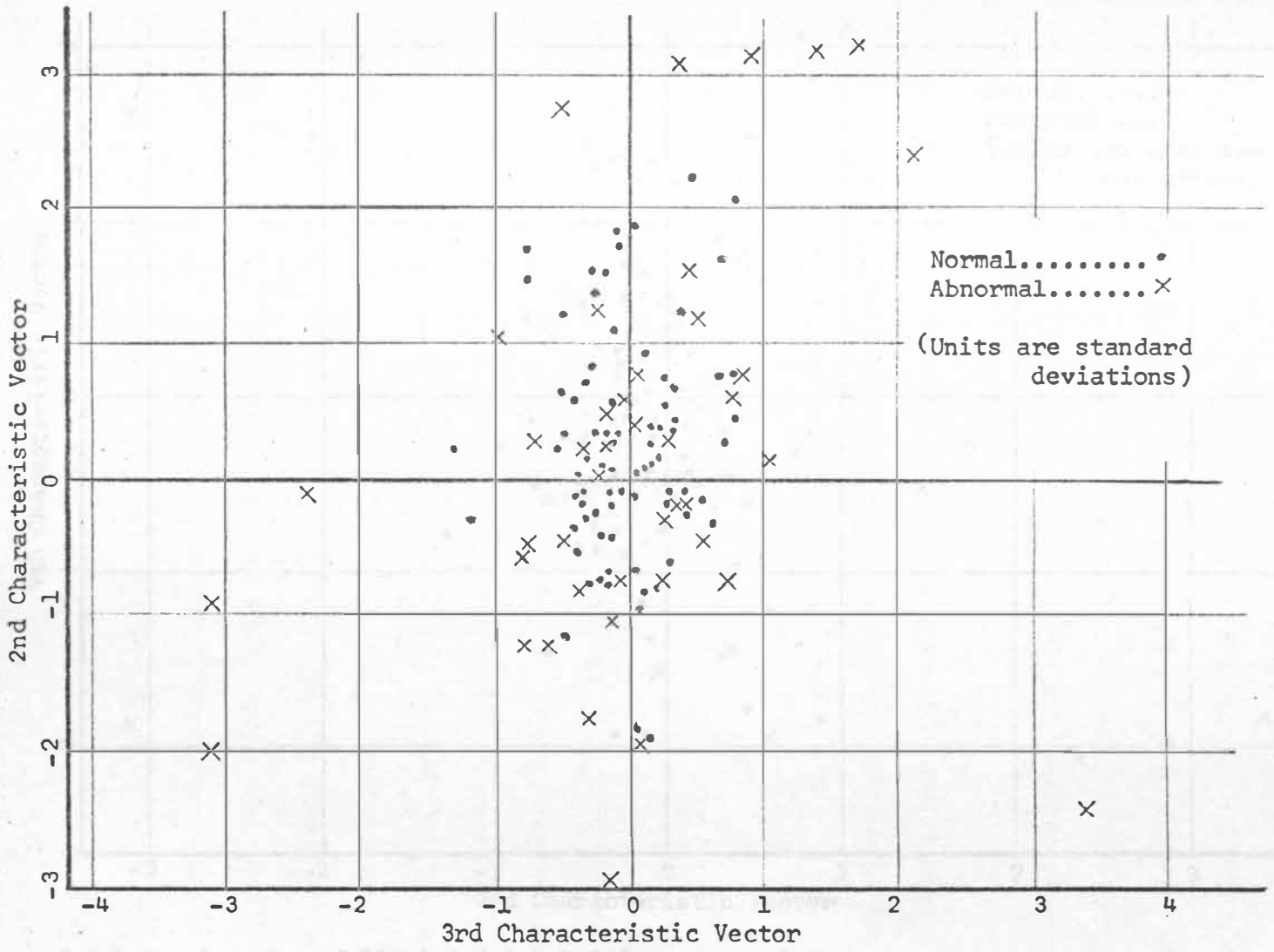
It was observed that the expected tendency for each type of vector to cluster about its average did not exist.

On examining the vector magnitudes of the ECGs one finds the average magnitude is much larger than the distance between the two average points. This accounts for the fact that the minimum-distance-to-the-mean criteria was not a good method for this type of categorization.

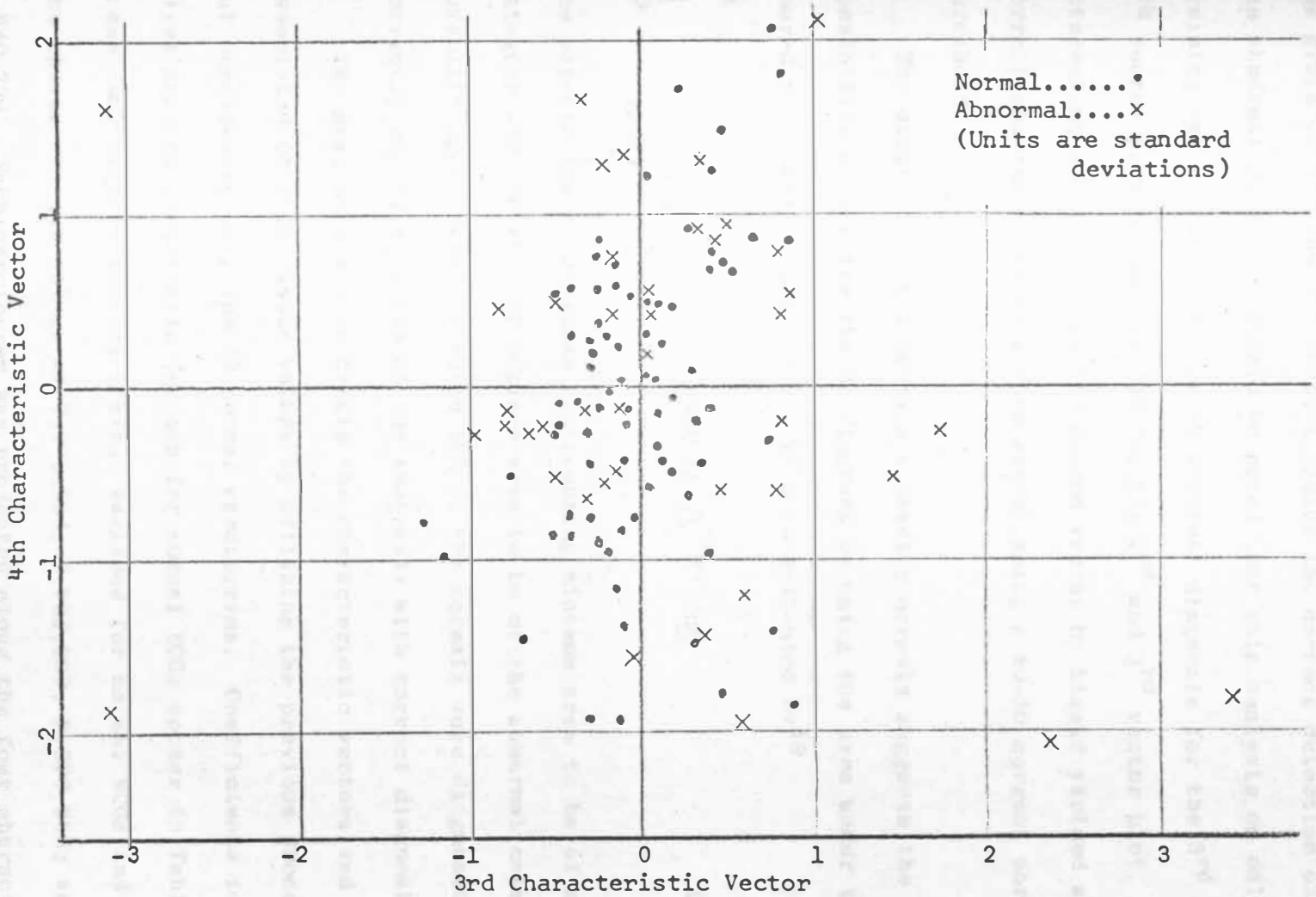
The next attempt was to decompose each cardiogram along the first four characteristic vectors. This produced the matrix $\underset{\sim}{X}$ composed of the four vector lengths. These vector lengths were converted into units of standard deviation. Plots of these results are shown in Graphs 2, 3, and 4. One observes a tendency for the abnormals to disperse more than the normals. This tendency was then capitalized upon in an attempt to categorize the ECGs. Drawing a circle about the center of the graph of the first two vectors such that 50% of the normals were enclosed by the circle yields 72% of the abnormals located outside the circle. A similar criterion applied to the graph of the 2^{nd} and 3^{rd} vectors yields correct abnormal detection of 77%; however,



Graph 2. Lengths of ECGs' 1st & 2nd 75% Normal-25% Abnormal Characteristic Vectors



Graph 3. Lengths of ECGs' 2nd & 3rd 75% Normal-25% Abnormal Characteristic Vectors



Graph 4. Lengths of ECGs' 3rd & 4th 75% Normal-25% Abnormal Characteristic Vectors

the graph of 3rd and 4th vectors yields the correct detection on 79% of the abnormal ECGs. It should be noted that this analysis on only the training set yielded 96% correct abnormal diagnosis for the 3rd and 4th vector plot on down to 76% for the 2nd and 3rd vector plot. An interesting point was that the second vector by itself yielded a 70% correct abnormal detection when establishing a 50-50 correct normal threshold.

The dispersion of abnormals exceeding normals suggests the possibility of sorting the cardiograms by using the area under the four-dimensional normal curve. This is evaluated by[19]

$$\int_0^X \frac{1}{(2\pi)^{k/2} |\underset{\sim}{\Lambda}_x|^{\frac{1}{2}}} \exp(-\tfrac{1}{2} \underset{\sim}{X} \underset{\sim}{\Lambda}_x^{-1} \underset{\sim}{X}') \, d\underset{\sim}{X} \qquad (31)$$

One expects the cardiograms representing minimum area to be of normal category and the ones of maximum area to be of the abnormal category. Establishing a threshold where 50% of the normals were diagnosed correctly resulted in 70% of the abnormals with correct diagnosis.

The next step was to create the characteristic vectors and associated characteristic values by utilizing the previous process but considering only the 32 normal cardiograms. Coefficients for the first four characteristic vectors for normal ECGs appear in Table 4. These four vectors possess maximum variance for normal ECGs and possess the characteristic values of 7,975,564; 3,206,069; 2,920,663; and 1,540,739. Each cardiogram was projected along the four characteristic

TABLE 4

COEFFICIENTS OF CHARACTERISTIC VECTORS 1, 2, 3, AND 4
OBTAINED FROM NORMAL COVARIANCE MATRIX

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
1	-.031	.025	.062	-.094
2	-.037	.048	.026	-.042
3	-.040	.040	.036	-.038
4	-.041	.063	.023	-.034
5	-.046	.044	.041	-.047
6	-.038	.056	.021	-.025
7	-.042	.044	.038	-.035
8	-.038	.041	.018	-.030
9	-.042	.059	.030	-.035
10	-.042	.068	.028	-.024
11	-.049	.094	.041	-.043
12	-.037	.101	.052	-.028
13	-.039	.081	.047	-.042
14	-.034	.047	.027	-.035
15	-.036	.056	.026	-.040
16	-.037	.026	.017	-.034
17	-.040	.008	.011	-.019
18	-.037	.014	.030	-.002
19	-.050	.008	.041	.001
20	-.052	.001	.056	-.023

TABLE 4 (cont.)

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
21	-.058	.010	.066	-.058
22	-.060	.003	.089	-.069
23	-.062	-.004	.085	-.056
24	-.053	-.006	.097	-.019
25	-.046	-.014	.078	-.004
26	-.045	-.023	.100	.005
27	-.044	-.029	.097	.007
28	-.049	-.007	.119	.004
29	-.067	.004	.127	-.002
30	-.031	-.130	.228	-.104
31	.137	-.296	.351	-.152
32	.404	-.374	.369	.010
33	.600	-.039	.120	.157
34	.498	.419	-.134	.132
35	.157	.552	.095	-.084
36	-.034	.318	.218	-.078
37	-.045	.044	.177	.074
38	-.066	-.084	.075	.264
39	-.070	-.113	-.004	.428
40	-.073	-.073	-.011	.353
41	-.052	-.003	.004	.029
42	-.055	.026	.014	-.061

TABLE 4 (cont.)

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
43	-.057	.008	.008	-.035
44	-.057	-.008	-.010	-.027
45	-.051	-.004	-.004	-.023
46	-.057	-.003	-.015	-.018
47	-.049	-.008	-.014	-.027
48	-.048	.019	-.018	-.029
49	-.055	-.026	-.022	-.027
50	-.035	.005	-.026	-.031
51	-.037	-.002	-.030	-.043
52	-.047	-.054	-.069	-.058
53	-.041	-.030	-.080	-.084
54	-.043	-.055	-.108	-.083
55	-.016	-.059	-.089	-.104
56	.015	-.048	-.099	-.106
57	.019	-.085	-.107	-.134
58	.046	-.082	-.146	-.141
59	.062	-.080	-.162	-.156
60	.087	-.097	-.188	-.154
61	.100	-.070	-.200	-.159
62	.110	-.076	-.218	-.132
63	.097	-.087	-.220	-.096
64	.095	-.080	-.217	-.064

TABLE 4 (cont.)

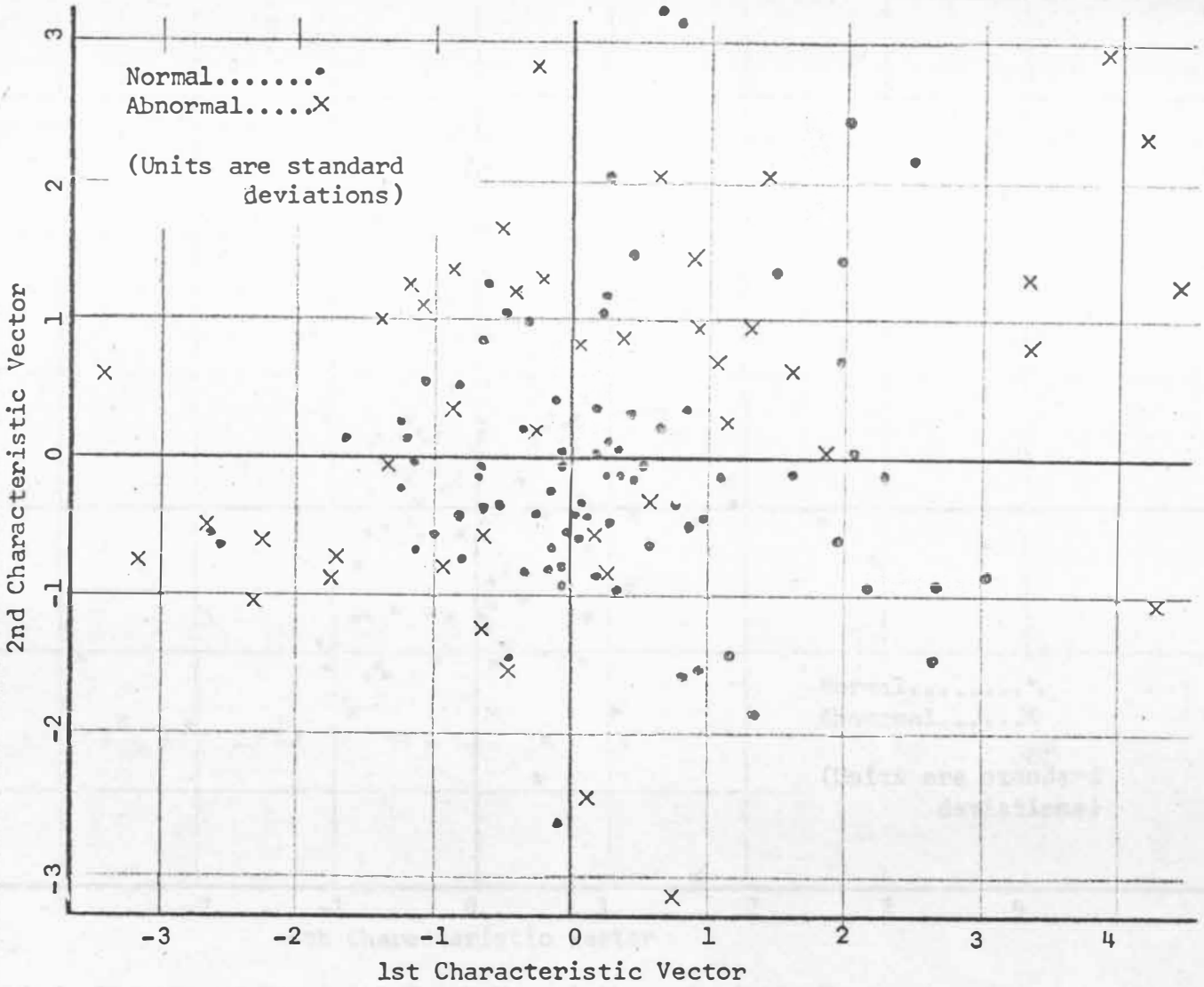
Coeff. No.	Vector Number. 1st	2nd	3rd	4th
65	.067	-.076	-.195	-.009
66	.054	-.065	-.171	.053
67	.029	-.031	-.142	.115
68	-.004	-.061	-.132	.169
69	-.036	-.028	-.118	.225
70	-.037	-.004	-.079	.236
71	-.037	-.004	-.051	.196
72	-.027	.028	-.026	.180
73	-.039	.021	-.035	.173
74	-.042	.007	-.022	.160
75	-.045	.030	-.025	.123

vectors with these four values forming the matrix $\underset{\sim}{X}$. When plotting ECGs by these vectors one would expect the tendency for the abnormal ECGs to cluster in the center. It was discovered that the opposite occurs. The area under the normal curve was now evaluated by (31). Forty-four abnormal and seventy-eight normal ECGs (including the 32 utilized as the training set) were analyzed. The decision threshold level was set so 50% of the normals were diagnosed as normal. Eighty percent of the abnormals were then diagnosed as abnormal. Adjusting the threshold such that 50% of the abnormals were diagnosed correctly yielded 77% of the normals diagnosed as normal.

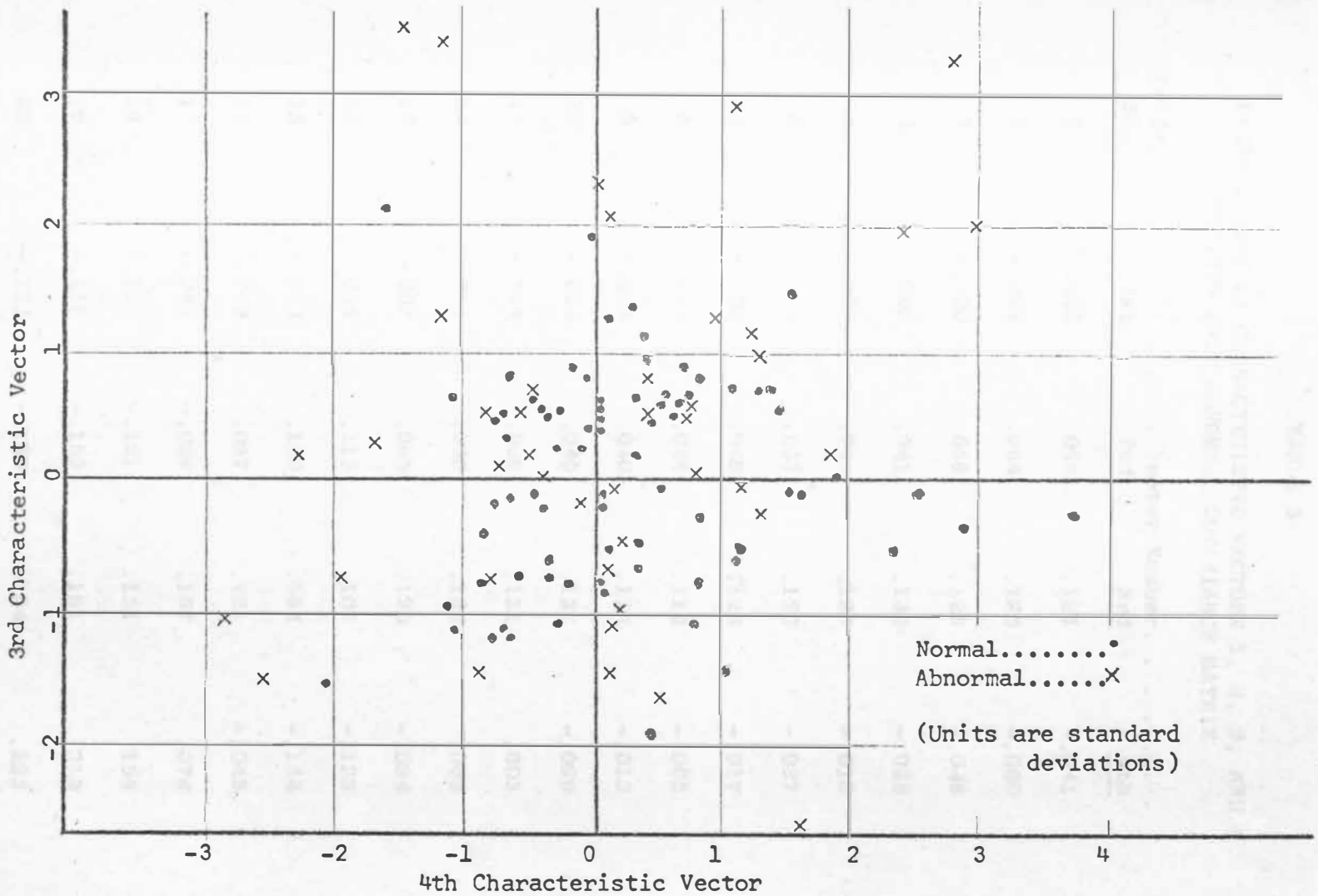
Adding the first two vectors (in units of their standard deviations) to give one vector, basing the sorting on this vector's length, and setting the threshold such that 50% of the normals were diagnosed correctly yielded 82% of the abnormals correctly diagnosed. This can be observed on Graph 5 by drawing a circle about the center with a radius of approximately one standard deviation. Setting this threshold such that 50% of the abnormals were diagnosed correctly yielded 72% of the normals properly diagnosed.

Using vectors three and four as the separation vectors, as in Graph 6, shows that setting the threshold at the usual 50% correct normal diagnostic level yields 70% of the abnormals diagnosed correctly.

The strong tendency for abnormal ECGs to disperse more than normal ECGs suggests that the characteristic vectors should be based on abnormal data. This was done by utilizing 28 abnormal ECGS as the



Graph 5. Lengths of ECGs' 1st & 2nd Normal Characteristic Vectors



Graph 6. Lengths of ECGs' 3rd & 4th Normal Characteristic Vectors

TABLE 5

COEFFICIENTS OF CHARACTERISTIC VECTORS 1, 2, 3, AND 4
OBTAINED FROM ABNORMAL COVARIANCE MATRIX

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
1	.002	.064	.137	-.041
2	-.002	.054	.125	-.050
3	-.000	.056	.129	-.046
4	-.005	.041	.132	-.025
5	-.011	.043	.126	-.016
6	-.001	.058	.120	-.037
7	-.005	.048	.123	-.017
8	-.013	.048	.114	-.002
9	-.004	.040	.114	-.013
10	-.008	.040	.127	-.029
11	-.014	.009	.131	.003
12	-.011	.000	.124	.005
13	.007	.045	.120	-.034
14	.048	.113	.106	-.122
15	.073	.150	.084	-.154
16	.015	.057	.121	-.042
17	-.047	-.039	.137	.074
18	-.105	-.121	.151	.158
19	-.116	-.153	.161	.219
20	-.135	-.180	.154	.225

TABLE 5 (cont.)

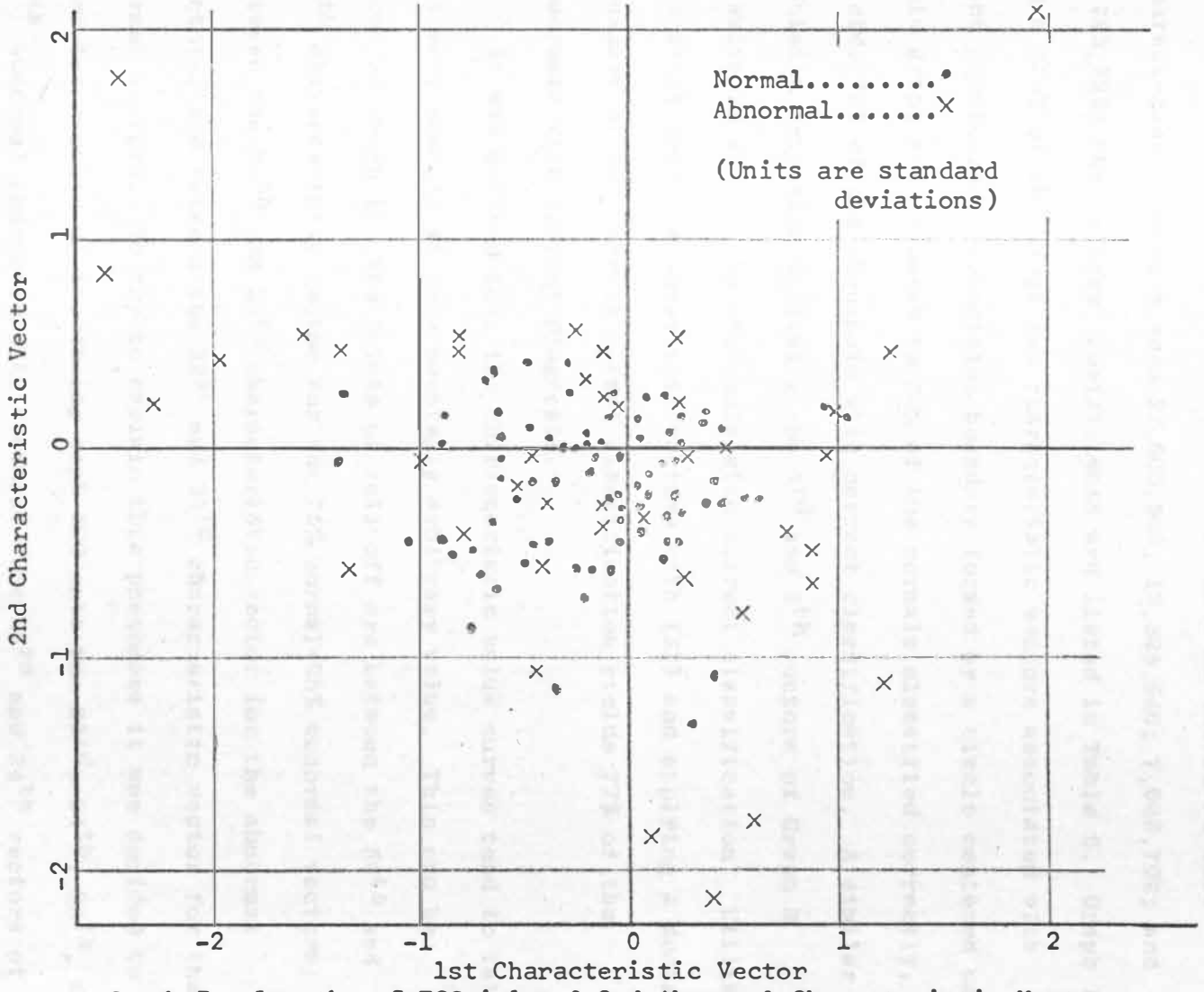
Coeff. No.	Vector Number. 1st	2nd	3rd	4th
21	-.131	-.184	.151	.224
22	-.116	-.158	.140	.204
23	-.097	-.149	.136	.159
24	-.086	-.117	.134	.138
25	-.048	-.062	.099	.076
26	-.013	-.022	.080	.019
27	.002	.024	.058	.005
28	.008	.031	.071	-.018
29	-.003	.037	.096	-.057
30	-.037	.109	.139	-.130
31	-.147	.142	.148	-.254
32	-.349	.112	.033	-.366
33	-.519	.059	-.170	-.238
34	-.473	.032	-.325	.032
35	-.240	.137	-.307	.176
36	-.002	.195	-.191	.263
37	.034	.221	-.110	.237
38	.039	.242	-.085	.206
39	.048	.223	-.071	.177
40	.040	.193	-.085	.172
41	.049	.196	-.076	.149
42	.056	.145	-.080	.152

TABLE 5 (cont.)

Coeff. No.	Vector Number. 1st	2nd	3rd	4th
43	.066	.131	-.062	.069
44	.072	.098	-.044	.019
45	.086	.074	-.027	-.055
46	.089	.067	-.024	-.055
47	.103	.057	-.019	-.069
48	.104	.044	-.025	-.066
49	.103	.046	-.026	-.063
50	.109	.051	-.033	-.090
51	.110	.035	-.030	-.069
52	.117	.032	-.031	-.079
53	.118	.022	-.032	-.065
54	.103	.023	-.041	-.052
55	.095	.002	-.054	-.061
56	.093	-.002	-.065	-.061
57	.071	-.041	-.070	-.046
58	.066	-.062	-.073	-.048
59	.063	-.090	-.094	-.061
60	.045	-.118	-.085	-.027
61	.049	-.118	-.097	-.058
62	.037	-.148	-.108	-.040
63	.038	-.162	-.123	-.047

TABLE 5 (cont.)

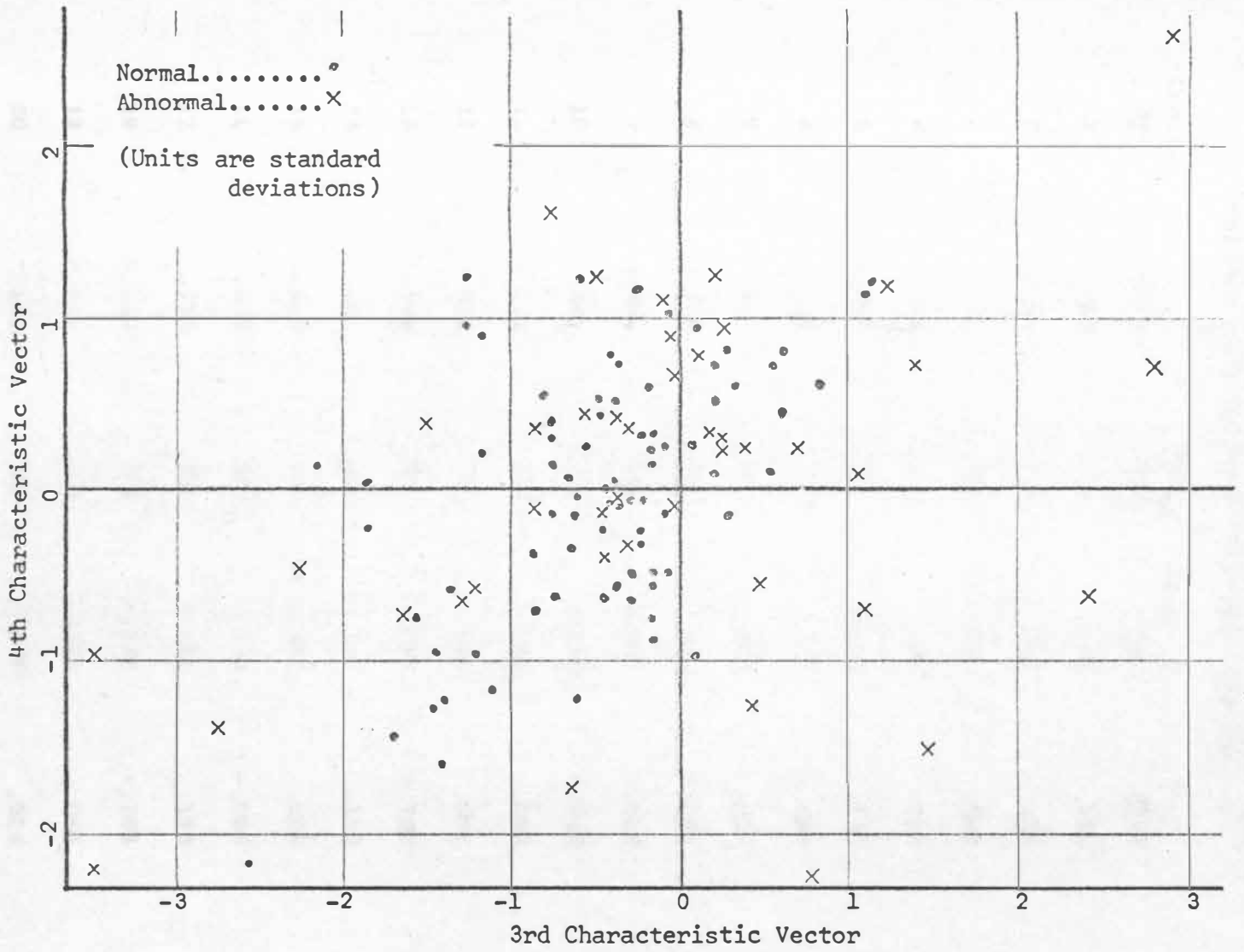
Coeff. No.	Vector Number. 1st	2nd	3rd	4th
64	.029	-.167	-.121	-.074
65	.032	-.155	-.128	-.053
66	.029	-.171	-.123	-.064
67	.031	-.177	-.138	-.079
68	.044	-.180	-.130	-.044
69	.056	-.178	-.126	-.014
70	.044	-.168	-.084	-.005
71	.058	-.135	-.070	-.009
72	.070	-.112	-.066	-.027
73	.054	-.097	-.057	.000
74	.059	-.102	-.052	.003
75	.075	-.079	-.061	.009



Graph 7. Lengths of ECGs' 1st & 2nd Abnormal Characteristic Vectors

training set. The characteristic vector and the characteristic value matrices were formed. Characteristic values of the first four characteristic vectors were 27,500,592; 12,525,646; 7,088,702; and 5,793,733; the vectors' coefficients are listed in Table 5. Graph 7 is a plot of the first two characteristic vectors associated with each cardiogram. A decision boundary formed by a circle centered on this graph and adjusted to 50% of the normals classified correctly, yields 79% of the abnormals with correct classification. A similar decision criterion utilizing the 3^{rd} and 4^{th} vectors of Graph 8 results in 65% of the abnormals with correct classification. Utilizing the first four characteristic vectors with (31) and applying a decision boundary of 50% correct normal classification yields 77% of the abnormals with correct diagnosis.

It was noticed that the characteristic value curves tend to fall off very sharply at some seemingly arbitrary value. This can be noted on Graph 1. The points of fall-off are between the 58^{th} and 59^{th} characteristic vector for the 75% normal-25% abnormal vectors; between the 24^{th} and 25^{th} characteristic vector for the abnormal vectors; and between the 32^{nd} and 33^{rd} characteristic vector for the normal vectors. To try to explain this phenomena it was decided to plot the results of impressing each ECG onto the 23^{rd}, 24^{th}, 25^{th}, and 26^{th} abnormal characteristic vectors. The 23^{rd} and 24^{th} vectors of each ECG are plotted on Graph 9. Due to the small variances of 9,990 and 6,862, there is the expected tendency for the training set ECGs



Graph 8. Lengths of ECGs' 3rd & 4th Abnormal Characteristic Vectors

4th

TABLE 6

COEFFICIENTS OF CHARACTERISTIC VECTORS 23, 24, 25, AND 26
OBTAINED FROM ABNORMAL COVARIANCE MATRIX

Coeff. No.	Vector Number. 23rd	24th	25th	26th
1	.032	.136	-.082	.124
2	-.084	-.145	-.082	.111
3	.113	.086	-.114	.093
4	-.034	-.107	-.135	.114
5	-.055	.184	-.172	.145
6	-.172	-.025	-.119	-.109
7	.027	.043	-.086	.044
8	-.007	-.119	-.062	-.163
9	-.055	.192	-.081	-.014
10	-.181	.035	-.152	-.049
11	-.171	.168	-.093	.029
12	-.003	-.192	-.133	.031
13	.188	-.002	-.137	-.150
14	.191	-.131	-.117	.110
15	-.016	.042	-.066	.026
16	.089	-.085	-.143	-.050
17	.058	.028	-.116	.175
18	-.135	-.279	-.152	-.106
19	-.104	-.105	-.111	.047
20	.142	.166	-.077	.014

TABLE 6 (cont.)

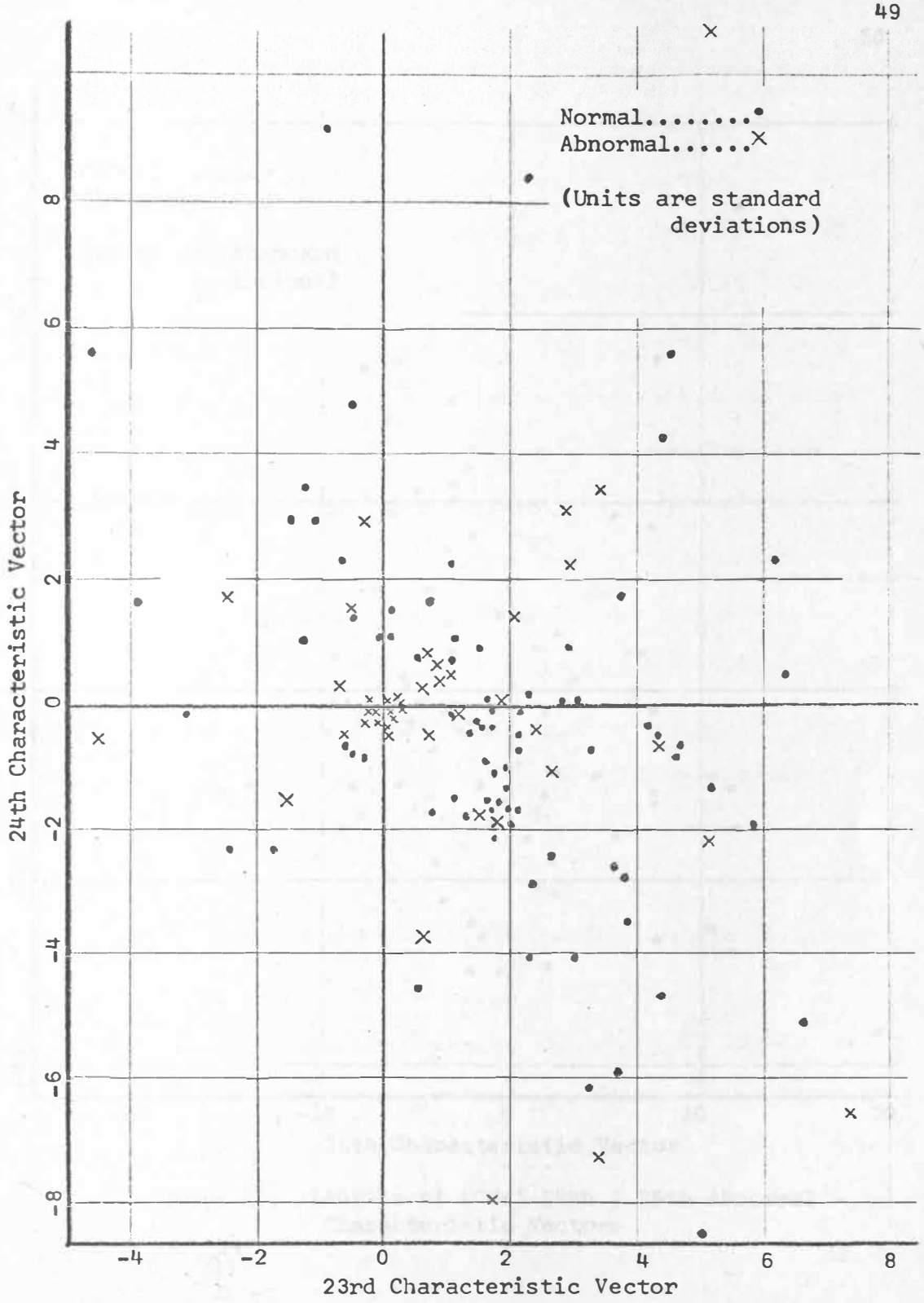
Coeff. No.	Vector Number. 23rd	24th	25th	26th
21	.084	.130	-.132	-.269
22	-.040	.123	-.095	-.083
23	-.015	-.046	-.091	.253
24	.008	-.056	-.073	.086
25	-.145	-.147	-.185	-.090
26	.091	.105	-.057	-.002
27	.057	-.048	-.058	-.239
28	.221	.020	-.108	-.121
29	.037	.099	-.166	.051
30	-.114	-.051	-.098	.023
31	.053	-.083	-.072	-.142
32	.159	.072	-.146	.150
33	-.231	.051	-.090	-.026
34	.165	-.129	-.143	.077
35	-.035	.072	-.065	-.132
36	-.199	.104	-.137	.080
37	.162	-.165	-.111	.065
38	.147	.020	-.129	-.084
39	-.118	-.097	-.128	-.079
40	.106	-.111	-.028	-.018
41	.015	.117	-.145	-.123
42	.113	.164	-.155	.177

TABLE 6 (cont.)

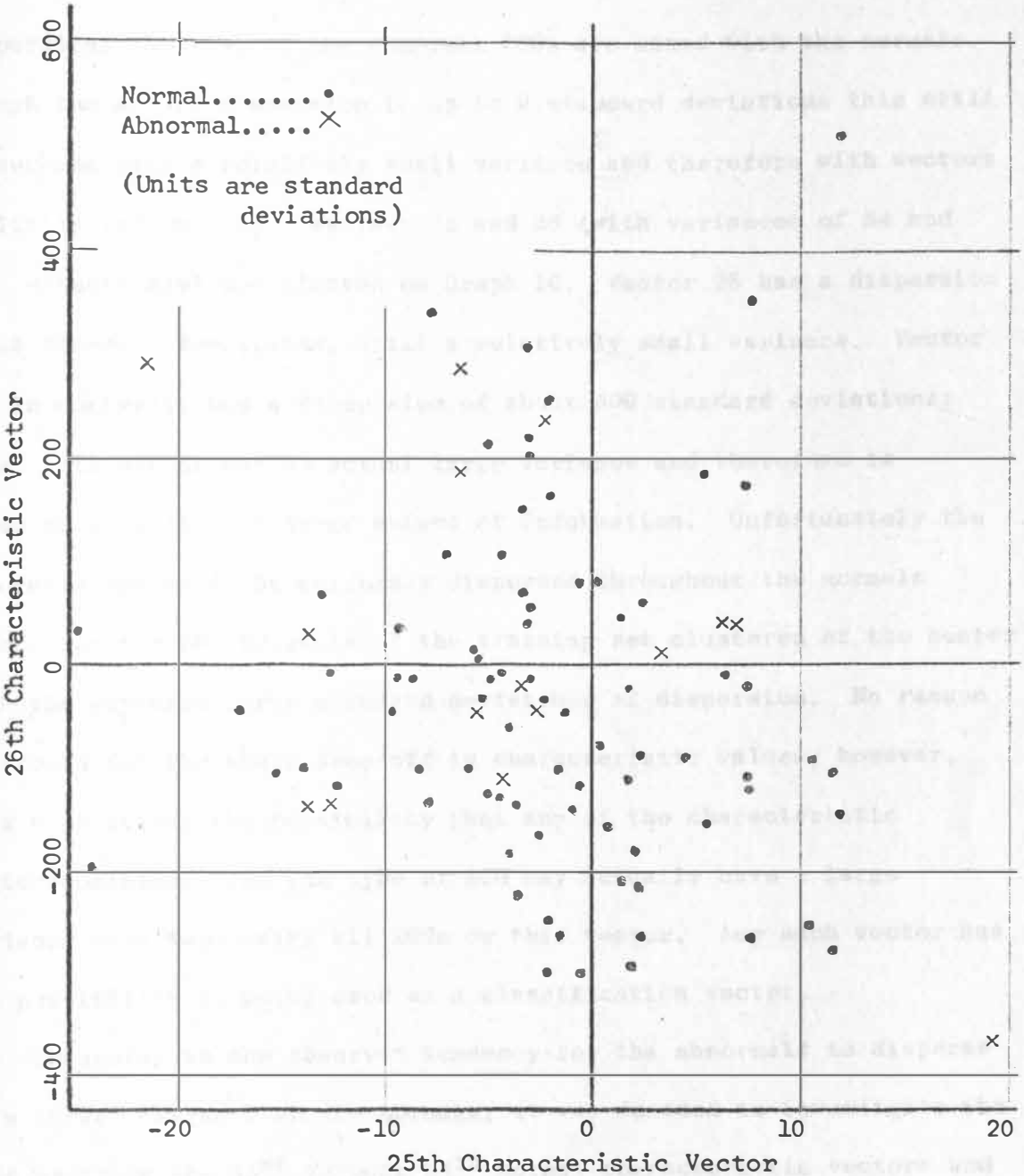
Coeff. No.	Vector Number. 23rd	24th	25th	26th
43	-.122	.014	-.073	.271
44	-.183	-.070	-.092	.159
45	-.128	.171	-.167	-.072
46	-.183	-.053	-.134	-.013
47	.042	-.168	-.152	-.141
48	.196	.150	-.105	-.068
49	.009	-.026	-.028	-.055
50	-.003	.156	-.158	-.214
51	.033	.001	-.098	.074
52	.017	-.123	-.161	-.208
53	-.016	-.070	-.072	.143
54	.016	-.024	-.107	-.083
55	-.105	-.057	-.113	.127
56	-.030	.097	-.099	.089
57	.008	.104	-.069	.121
58	-.121	-.193	-.089	-.047
59	-.146	.137	-.122	-.128
60	.176	.020	-.129	-.011
61	.155	-.195	-.118	.031
62	-.004	.191	-.121	-.041
63	.062	-.044	-.100	-.039
64	.109	-.084	-.094	.019

TABLE 6 (cont.)

Coeff. No.	Vector Number. 23rd	24th	25th	26th
65	.057	-.019	-.124	-.063
66	-.089	.013	-.141	-.014
67	-.172	-.099	-.079	-.129
68	-.136	.108	-.128	-.062
69	.091	-.029	-.093	.069
70	.215	-.029	-.165	.145
71	.014	-.102	-.065	.159
72	-.035	.208	-.100	.061
73	.111	.108	-.104	-.042
74	-.133	.053	-.069	.137
75	-.049	-.183	-.157	.125



Graph 9. Lengths of ECGs' 23rd & 24th Abnormal Characteristic Vectors



Graph 10. Lengths of ECGs' 25th & 26th Abnormal Characteristic Vectors

to cluster in the center with the usual three standard deviations of dispersion; the rest of the abnormal ECGs are mixed with the normals. Though the actual dispersion is up to 8 standard deviations this still leaves one with a relatively small variance and therefore with vectors of little information. Vectors 25 and 26 (with variances of 54 and 4.7 respectively) are plotted on Graph 10. Vector 25 has a dispersion of 25 standard deviations, still a relatively small variance. Vector 26, in contrast, has a dispersion of about 500 standard deviations; thus, this vector has an actual large variance and therefore is expected to contain a large amount of information. Unfortunately the abnormals appear to be uniformly dispersed throughout the normals except for the 28 abnormals of the training set clustered at the center with the expected three standard deviations of dispersion. No reason was found for the sharp drop-off in characteristic values; however, this does stress the possibility that any of the characteristic vectors obtained from one type of ECG may actually have a large variance when impressing all ECGs on this vector. Any such vector has the possibility of being used as a classification vector.

Returning to the observed tendency for the abnormals to disperse by a larger degree than the normals, it was decided to investigate the ECGs by using the 31^{st} through 34^{th} normal characteristic vectors and check if any classification could be done with these. It was discovered that the 31^{st} and 32^{nd} vectors had a very small dispersion (and therefore little information.) Vectors 33 and 34 are plotted on Graph 11.

TABLE 7

COEFFICIENTS OF CHARACTERISTIC VECTORS 31, 32, 33, AND 34 OBTAINED FROM NORMAL COVARIANCE MATRIX

Coeff. No.	Vector Number. 31st	32nd	33rd	34th
1	-.012	-.065	-.007	.018
2	.058	.221	.181	.081
3	-.115	-.180	-.018	-.074
4	-.029	-.006	.062	.043
5	-.183	.032	-.196	.084
6	-.161	-.041	.014	.083
7	.148	.033	-.050	.043
8	.125	.032	-.013	.027
9	.317	-.152	-.001	-.064
10	-.001	-.111	-.041	-.154
11	-.010	-.106	-.013	-.018
12	-.029	-.019	-.081	.226
13	.030	.205	.169	-.083
14	-.143	.232	-.064	-.013
15	-.182	-.057	-.175	-.070
16	-.004	-.018	-.072	.111
17	-.133	-.174	-.141	-.070
18	.140	.276	.150	.180
19	.281	-.185	.123	-.073
20	-.177	-.202	.204	-.017

TABLE 7 (cont.)

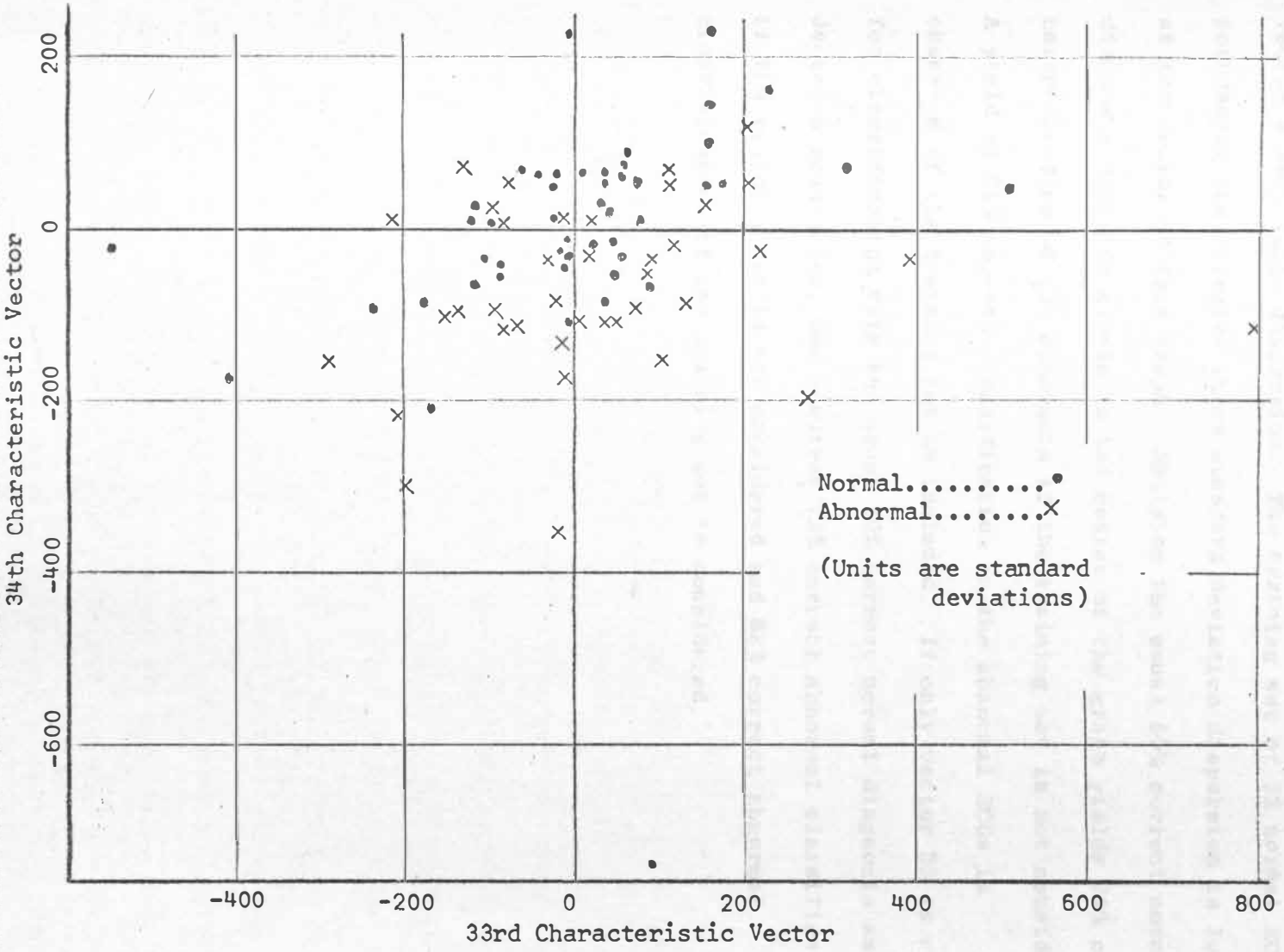
Coeff. No.	Vector Number. 31st	32nd	33rd	34th
21	-.116	-.058	-.088	.128
22	-.071	.148	.093	-.012
23	.274	-.051	-.082	.095
24	.082	.106	-.046	-.003
25	.028	.046	-.107	.214
26	-.139	.011	.102	.092
27	-.074	.073	-.047	.035
28	.106	-.028	.050	-.258
29	-.035	-.106	-.083	-.068
30	.131	-.056	-.088	.225
31	-.046	.001	.205	-.130
32	-.020	-.006	-.188	.080
33	.032	-.001	.080	.003
34	-.006	.014	-.108	.006
35	-.019	.008	.109	.026
36	.084	-.030	-.177	-.008
37	-.045	.119	.009	-.035
38	-.132	-.067	.110	.130
39	.099	-.074	-.045	.024
40	.005	.135	-.150	-.125
41	-.054	.077	.058	.206
42	-.094	.007	-.060	.119

TABLE 7 (cont.)

Coeff. No.	Vector Number. 31st	32nd	33rd	34th
43	-.117	.173	.110	-.015
44	.149	-.099	.129	-.093
45	.049	-.016	.078	.053
46	.138	.124	-.024	.011
47	.025	-.058	.006	.043
48	-.061	-.269	-.044	.052
49	.013	.023	-.053	-.078
50	.122	.182	-.180	.153
51	-.104	.013	-.097	-.027
52	.147	.002	-.270	-.170
53	.014	-.087	-.065	.094
54	.041	.171	.038	.067
55	-.122	-.076	-.178	-.188
56	-.205	.192	.087	-.017
57	-.069	.137	.071	-.129
58	-.136	-.071	.070	.006
59	.093	-.111	-.103	.335
60	.013	.037	-.041	.058
61	.144	-.012	.141	-.096
62	.093	.006	-.162	.100
63	.061	-.008	.225	.124
64	-.074	-.157	.010	-.181

TABLE 7 (cont.)

Coeff. No.	Vector Number. 31st	32nd	33rd	34th
65	-.028	.038	-.173	-.128
66	-.073	-.049	-.009	.299
67	-.052	-.024	.182	.040
68	.157	.253	-.141	-.205
69	.037	-.131	.051	.093
70	-.075	-.103	.019	.115
71	-.169	.160	-.159	.005
72	.151	.031	.099	.023
73	-.110	.040	-.016	.002
74	.146	-.151	.037	.095
75	-.105	-.130	.234	-.003



Graph 11. Lengths of ECGs' 33rd & 34th Normal Characteristic Vectors

It is noted that the 34th vector has a modest dispersion and the 33rd vector a very large dispersion. The training set of 32 normal ECGs possessing the expected three standard deviation dispersion is located at the center of this graph. Applying the usual 50% correct normal diagnosis decision circle to the center of the graph yields 79% correct categorization of the abnormals if the training set is not considered. A yield of 91% correct classification on the abnormal ECGs is observed if the training set is included. If only vector 33 is used for classification with the usual 50% correct normal diagnosis as the decision boundaries, one acquires 72% correct abnormal classification if the training set is not considered and 88% correct abnormal classification if the training set is considered.

CONCLUSIONS

Several problems presented themselves during the heartbeat detection portion of the computer program. Steinberg[18] pointed out that the heartbeat interval was always greater than 0.64 seconds. This amounts to 94 beats per minute. Pulse rates are often greater. If one wishes to consider these higher pulse rates then a smaller interval than 160 data points must be used in which to search for points of maximum negative slope.

Another assumption used in the heartbeat detection program was that the maximum negative slope of each heartbeat never varies by over 30%[2,18]. During the heart rate evaluations of this data there were numerous occasions (approximately 40% of the ECGs) when the points of maximum negative slope varied by over 30%. This may have been due to the presence of noise on the line over which the cardiogram was sent; however, any cardiogram with an RMS noise greater than 50 millivolts was not considered. Others[2,8,18] have apparently used a similar technique for heartbeat detection (see Appendix B) but have not mentioned a similar difficulty. Another possible reason is that the sampling rate for recording the ECG was too slow. The rate used for recording the data for this thesis was 250 samples per second. The average time between the peaks of the R and S waves was .032 seconds or 8 data points at a sampling rate of 250 per second. Steinberg et al.[18] used data which was recorded at 625 samples per second.

Another problem was the shortage of samples. When using the 25% abnormal with 75% normal vectors, the 3rd and 4th vectors did the best job of classification. One expects the 1st and 2nd vectors to do the best classification when utilizing a training set composed of both categories. Further evidence of sample shortage was the fact that the percentage correct from the training set was usually noticeably different from the percentage correct from the samples outside the training set. The various percentages cited in this thesis should be considered an indication of magnitude and not an exact figure.

One of the significant observations of this thesis was the tendency for the abnormal ECGs to disperse more than the normals. This can be observed on Graph 1 for the first 20 characteristic vectors. Another area that displayed this tendency was the first characteristic vectors obtained from the normal training set of ECGs. These characteristic vectors are the vectors of maximum dispersion obtained from normal data. When abnormal ECGs were decomposed using these characteristic vectors (Graphs 5 and 6) the abnormals dispersed more than the normals. This phenomena might be explained if one considers the fact of the existence of various abnormal categories. One may consider the normals as a parent category and theorize that each abnormal category moves away from the normal category in a unique direction. The data for this thesis did not differentiate the abnormals into the various categories; therefore, this theory was not examined.

A most significant observation was that in categorizing the patterns the possibility exists of using characteristic vectors other than those associated with the largest variance. These characteristic vectors are those formed from a single category covariance matrix. This is not in opposition to the fact that a parameter's variance is proportional to its information content. Some of the vectors associated with a small variance from a single category acquired a very large variance when this measurement was obtained by considering both categories. This phenomenon was observed with the 26^{th} characteristic vector formed from abnormal data and the 33^{rd} characteristic vector formed from normal data. When the features of the 26^{th} abnormal characteristic vector were obtained for both categories, one observed the dispersion to be around 500 standard deviations; the 33^{rd} normal characteristic vector performed similarly.

The results obtained indicate the distinct possibility of sorting ECGs by vectors obtained from the Hotelling Procedure. This can be observed by examining Table 8. If sorting is to be between normal and possibly abnormal categories, the best procedure appears to be the forming of the covariance matrix from the normal data and then applying these vectors to a mixed group of ECGs. One experimentally selects those vectors that giving the best category separation and combines the vectors by some one process to give the final separation. Such a process could be by the radius method as used in this thesis. Probably, each vector should have a weighting factor based on its ability to separate the two categories.

TABLE 8

SUMMARY OF RESULTS

This table gives classification success of abnormal ECGs by various methods when the classification success of normal ECGs is at 50%.

Characteristic Vectors from 75% Normal-25% Abnormal Covariance Matrix

Procedure	% Success
Minimum distance to mean	no positive results
1st & 2nd characteristic vectors	72%
2nd & 3rd characteristic vectors	77%
3rd & 4th characteristic vectors	79%
2nd characteristic vector	70%
Equation (31)	70%

Characteristic Vectors from Abnormal Covariance Matrix

Procedure	% Success
1st & 2nd characteristic vectors	79%
3rd & 4th characteristic vectors	65%
Equation (31)	77%
23rd through 26th characteristic vectors	no positive results

Characteristic Vectors from Normal Covariance Matrix

Procedure	% Success
1st & 2nd characteristic vectors	82%
3rd & 4th characteristic vectors	70%
Equation (31)	80%
33rd & 34th characteristic vectors (without training set)	79%
33rd & 34th characteristic vectors (with training set)	91%
33rd characteristic vector (without training set)	72%
33rd characteristic vector (with training set)	88%

The reason for using the normal data as opposed to the abnormal data for the covariance matrix was the tendency for abnormal ECGs to disperse more readily than normal ECGs. When using abnormal data, the characteristic vectors of largest variance gave better separation than their normal counterparts. This suggests using the best vectors from both the normal and abnormal covariance matrices and utilizing the mixture to make the decision. If such a method is utilized it must be recognized that the vectors obtained from the one source are in general not orthogonal to those obtained from the other source. This would lead to a certain amount of redundancy between the two sets of vectors.

If separation is to be between the various categories of the abnormals, it is suggested that the covariance matrix be formed from data for each type of abnormality. It is expected that a single abnormality will behave in a fashion similar to the normal category of this thesis.

Following is a brief outline of further research that could be performed to improve the ECG categorization by the Hotelling process.

(1) Acquire more data including classification of the abnormal ECGs into the various abnormalities.

(2) Improve the heartbeat recognition procedure.

(3) Check each characteristic vector obtained from the covariance matrix of the desired data for its possibility of being a good candidate as a classification vector. Decide on which vectors are needed to implement classification at some predetermined level of confidence.

(4) Investigate the possibility of classification improvement by using data from other leads in conjunction with that from the X-lead.

(5) Investigate the possibility of using less measurements on the ECG and still maintain a predetermined level of confidence.

APPENDIX A

EIGENVALUE AND EIGENVECTOR DETERMINATION BY THE JACOBI METHOD[15]

Consider matrix $\underset{\sim}{A}$ as being real and symmetric. A set of eigenvectors $(\underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_n)$ and a set of eigenvalues $(\lambda_1,\ldots,\lambda_n)$ are associated with matrix $\underset{\sim}{A}$; thus,

$$\underset{\sim}{A}\underset{\sim}{a}_i = \lambda_i \underset{\sim}{a}_i \tag{32}$$

Let there exist an orthogonal matrix $\underset{\sim}{S}$ such that

$$\underset{\sim}{S}^T\underset{\sim}{A}\underset{\sim}{S} = \underset{\sim}{D} \tag{33}$$

$\underset{\sim}{D}$ is a diagonal matrix. One now has

$$D_{ik} = \epsilon_i \delta_{ik} \tag{34}$$

Premultiply (33) by $\underset{\sim}{S}$. This yields

$$\underset{\sim}{A}\underset{\sim}{S} = \underset{\sim}{S}\underset{\sim}{D} \tag{35}$$

A general term for this resultant matrix is

$$\sum_{k=1}^{n} A_{mk}S_{ki} = \sum_{k=1}^{n} S_{mk}D_{ki} = \sum_{k=1}^{n} S_{mk}\,\delta_{ik}\,\epsilon_i = S_{mi}\epsilon_i \tag{36}$$

A general term of (32) is

$$\sum_{k=1}^{n} A_{mk}a_{ki} = a_{ki}\lambda_i \tag{37}$$

It is observed by comparing (36) and (37) that the i^{th} column of $\underset{\sim}{S}$ can be equated with $\underset{\sim}{a}_i$ and the ϵ_i can be equated with λ_i. The problem now becomes that of finding an orthogonal matrix $\underset{\sim}{S}$ which will transform matrix $\underset{\sim}{A}$ into a diagonal matrix $\underset{\sim}{D}$ under a similarity

transformation. The matrix $\underset{\sim}{S}$ is then a matrix of the eigenvectors of $\underset{\sim}{A}$, and $\underset{\sim}{D}$ is a matrix of the associated eigenvalues of $\underset{\sim}{A}$.

Let $\underset{\sim}{S}$ be composed of a series of elementary orthogonal transformations where each transformation reduces one off-diagonal term (the pivotal element) to zero. Selection of the elementary transformation may be guided by the orthogonal transformation that rotates the X-Y axis in X-Y-Z space:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} \text{Cos}\theta & \text{Sin}\theta & 0 \\ -\text{Sin}\theta & \text{Cos}\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (38)$$

X' or Y' can be transformed to zero by the judicious selection of θ.

Define the orthogonal transformation (denoted by $\underset{\sim}{R}$) as

$$\begin{aligned} R_{pp} &= \text{Cos}\theta, & R_{qq} &= \text{Cos}\theta, \\ R_{pq} &= \text{Sin}\theta, & R_{qp} &= -\text{Sin}\theta, \qquad (39) \\ R_{ii} &= 1, \qquad R_{pk} = R_{iq} = R_{ik} = 0 \quad i,k \neq p,q \end{aligned}$$

Let matrix $\underset{\sim}{B}$ be formed by

$$\underset{\sim}{R}^T \underset{\sim}{A} \underset{\sim}{R} = \underset{\sim}{B} \qquad (40)$$

An arbitrary component of $\underset{\sim}{B}$ is

$$B_{il} = \sum_{j=1}^{n} \sum_{k=1}^{n} R_{ji} A_{jk} R_{kl} \qquad (41)$$

Using this equation the various elements of B become

$$\left.\begin{aligned} B_{pk} &= A_{pk}\text{Cos}\theta - A_{qk}\text{Sin}\theta \qquad (42a) \\ B_{qk} &= A_{pk}\text{Sin}\theta + A_{qk}\text{Cos}\theta \\ B_{ip} &= A_{ip}\text{Cos}\theta - A_{iq}\text{Sin}\theta \qquad i \neq p,q \\ B_{iq} &= A_{ip}\text{Sin}\theta + A_{iq}\text{Cos}\theta \qquad k \neq p,q \qquad (42b) \\ B_{ik} &= A_{ik} \qquad (42c) \end{aligned}\right\}$$

$$B_{pp} = A_{pp}Cos^2\theta + A_{qq}Sin^2\theta - 2A_{pq}Sin\theta Cos\theta \quad (43a)$$

$$B_{qq} = A_{pp}Sin^2\theta + A_{qq}Cos^2\theta + 2A_{pq}Sin\theta Cos\theta \quad (43b)$$

$$B_{pq} = (A_{pp} - A_{qq})Sin\theta Cos\theta + A_{pq}(Cos^2\theta - Sin^2\theta)$$

$$= \tfrac{1}{2}(A_{pp} - A_{qq})Sin2\theta + A_{pq}Cos2\theta \quad (44)$$

Off-diagonal element B_{pq} vanishes when

$$\tfrac{1}{2}(A_{pp} - A_{qq})Sin2\theta + A_{pq}Cos2\theta = 0 \quad (45)$$

Angle θ must be of such magnitude that

$$Tan2\theta = -\frac{2A_{pq}}{A_{pp}-A_{qq}} \quad (46)$$

A new matrix $\underset{\sim}{B}_i$ is formed for each off-diagonal element until all off-diagonal elements are reduced below some predetermined value. One now has

$$\underset{\sim}{D} = \underset{\sim}{B}_1\underset{\sim}{B}_2\cdots\underset{\sim}{B}_m \quad (47)$$

and

$$\underset{\sim}{S} = \underset{\sim}{R}_1\underset{\sim}{R}_2\cdots\underset{\sim}{R}_m \quad (48)$$

Jacobi initially showed that this method would converge by using the largest off-diagonal element as the pivotal element for each transformation. Others have shown that this method converges regardless of the order of pivotal element selection. The best method of pivotal selection for computer use seems to be that of selecting some desired level; examine each possible pivotal element by some regular sequence, and reduce all pivotal elements that are above this level. Then lower the level and repeat until $\underset{\sim}{D}$ is as close to diagonalization as desired.

To show that the Jacobi method converges one may square (42a) and (42b) to acquire

$$B_{pk}^2 + B_{qk}^2 = A_{pk}^2 + A_{qk}^2$$
$$B_{ip}^2 + B_{iq}^2 = A_{ip}^2 + A_{iq}^2 \qquad (49)$$

One notes by (42c) that the off-diagonal terms of $A_{ik}(i,k\neq p,q)$ are unaffected by the transformation. (49) shows the sum of squares of off-diagonal elements are invariant except for A_{pq}.

Utilizing (43a), (43b), and (44) one acquires

$$B_{pp}^2 + B_{pq}^2 + B_{qp}^2 + B_{qq}^2 = A_{pp}^2 + A_{qq}^2 + A_{pq}^2 + A_{qp}^2 \qquad (50)$$

The transformation has left $B_{pq}=0$ so one has

$$B_{pp}^2 + B_{qq}^2 = A_{pp}^2 + A_{qq}^2 + 2A_{pq}^2 \qquad (51)$$

$2A_{pq}^2$ has been removed from the sum of squares of off-diagonal terms and has been absorbed by the diagonal terms; thus, the process undergoes convergence.

APPENDIX B

HEARTBEAT DETECTION SUBPROGRAM

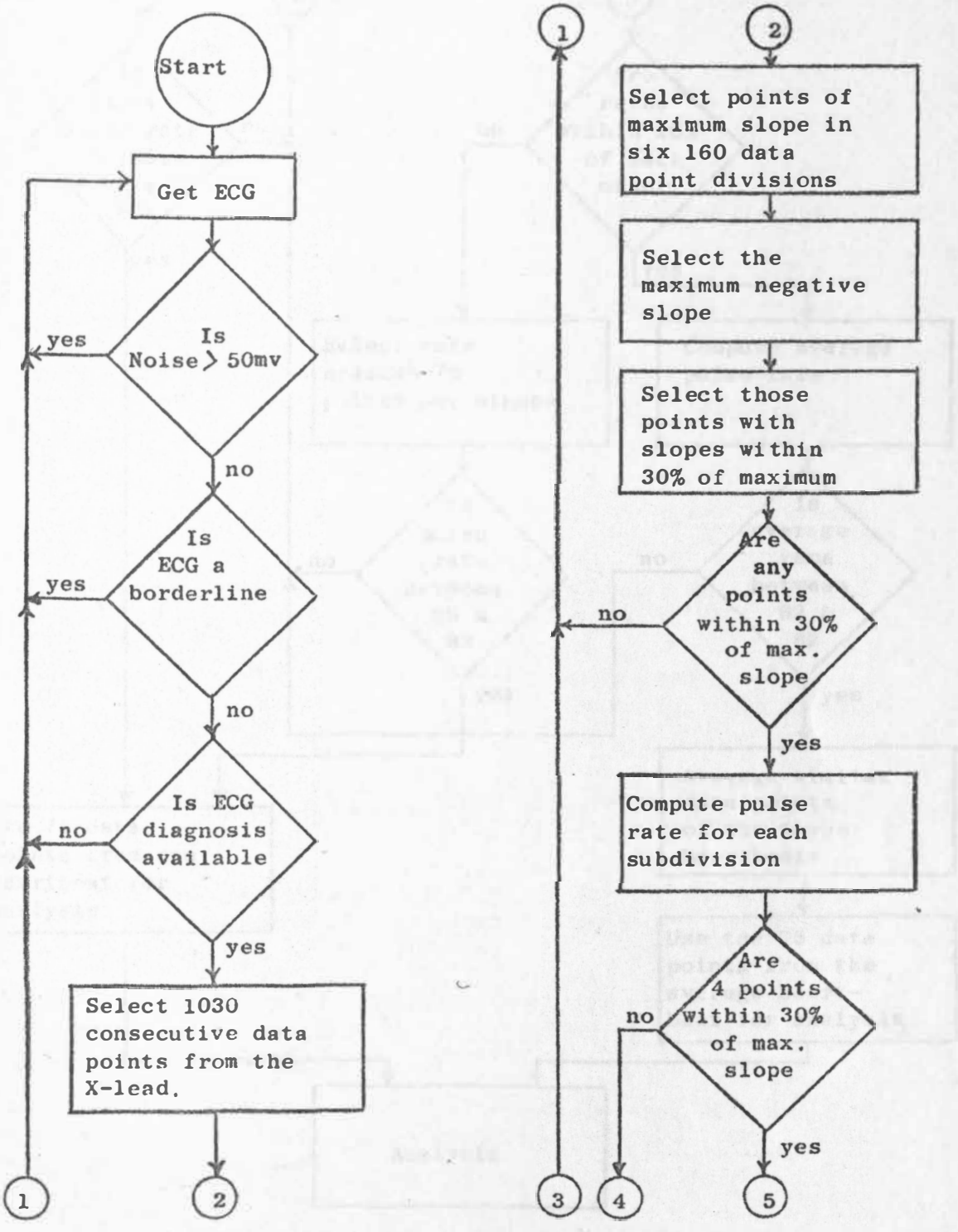


Figure 2. Heartbeat Detection Subprogram

HEARTBEAT DETECTION SUBPROGRAM (cont.)

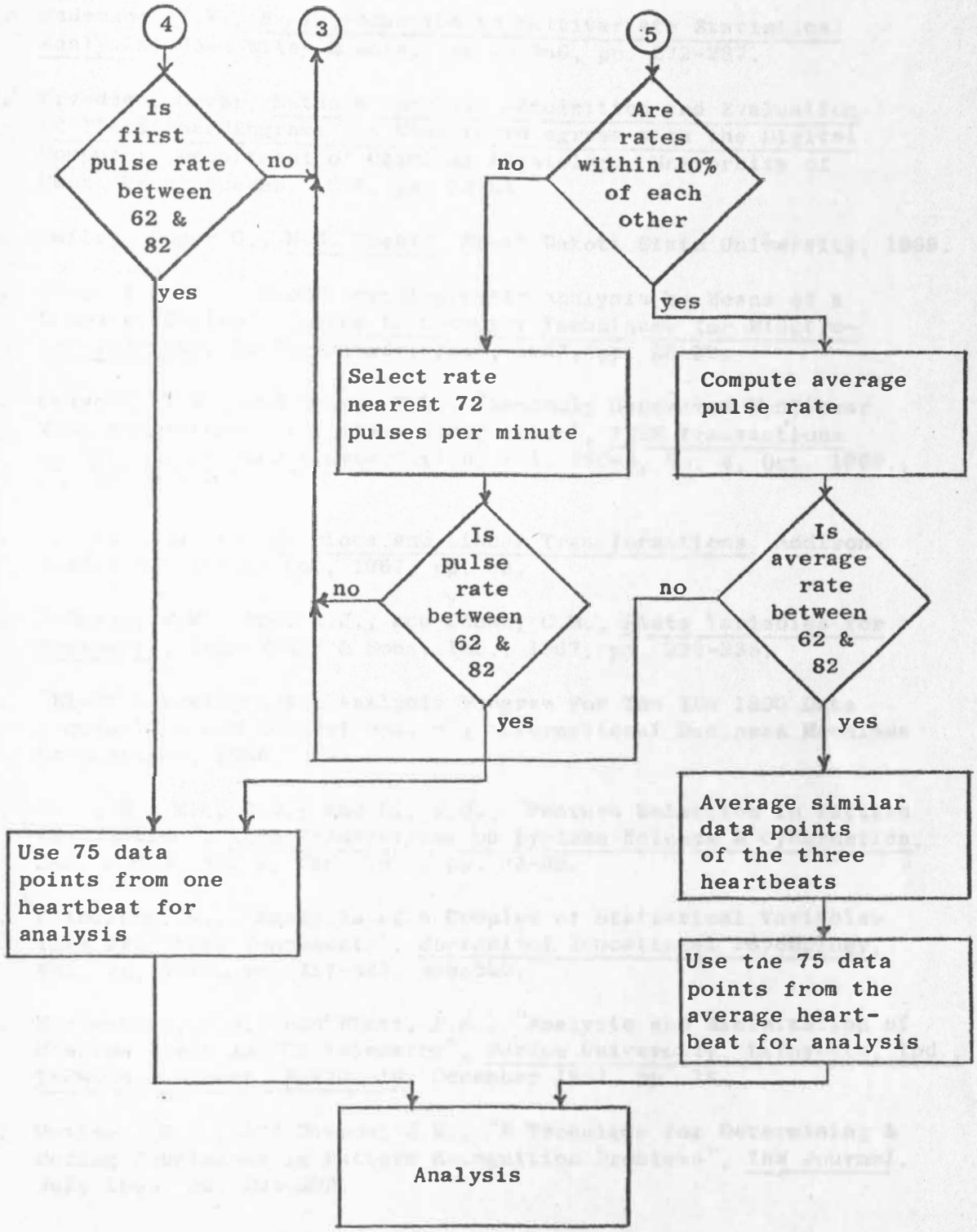


Figure 2 (cont.)

REFERENCES

1. Anderson, T.W., An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., 1966, pp. 272-287.

2. Arvedson, Orvar, Methods for Data Acquisition and Evaluation of Electrocardiograms and Vectorcardiograms with the Digital Computer, Department of Clinical Physiology, University of Umea, Umea, Sweden, 1968, pp. 18-54.

3. Bailey, Roger G., M.S. Thesis, South Dakota State University, 1969.

4. Caceres, C.A., "Electrocardiographic Analysis by Means of a Computer System", Topics in Computer Techniques for Electrocardiography, Cutler-Hammer, Inc., 1967, pp. 26-27.

5. Calvert, T.W., and Young, T.Y., "Randomly Generated Nonlinear Transformations for Pattern Recognition", IEEE Transactions on Systems Science & Cybernetics, Vol. SSC-5, No. 4, Oct. 1969., pp. 266-273.

6. Cullen, Charles, Matrices and Linear Transformations, Addison-Wesley Publishing Co., 1967, pp. 93.

7. DeRusso, P.M., Roy, R.J., and Close, C.M., State Variables for Engineers, John Wiley & Sons, Inc., 1967, pp. 232-235.

8. "Electrocardiographic Analysis Program For The IBM 1800 Data Acquisition and Control System", International Business Machines Corporation, 1968.

9. Fu, K.S., Min, P.J., and Li, T.J., "Feature Selection in Pattern Recognition", IEEE Transactions on Systems Science & Cybernetics, Vol. SSC-6, No. 1, Jan. 1970, pp. 33-39.

10. Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components", Journal of Educational Psychology, Vol. 24, 1933, pp. 417-441, 498-520.

11. Kurtenbach, A.J., and Wintz, P.A., "Analysis and Minimization of Message Error in PCM Telemetry", Purdue University, Lafayette, Ind., Technical Report TR-EE67-19, December 1967, pp. 18.

12. Mattson, R.L., and Dammon, J.E., "A Technique for Determining & Coding Subclasses in Pattern Recognition Problems", IBM Journal, July 1965, pp. 294-302.

13. Nagy, George, "State of the Art in Pattern Recognition", Proceedings of the IEEE, Vol. 56, No. 5, May 1968, pp. 836-862.

14. Peterson, D.W., and Mattson, R.L., "A Method of Finding Linear Discriminant Functions for a Class of Performance Criteria", IEEE Transactions on Information Theory, Vol. IT-12, No. 3, July 1966, pp. 380-387.

15. Ralston, A., and Wilf, H., "The Determination of the Characteristic Roots of a Matrix by the Jacobi Method", Mathematical Methods for Digital Computers, Vol. 1, John Wiley & Sons, Inc., 1960, pp. 84-91.

16. "Remote Multispectral Sensing in Agriculture", Laboratory for Agricultural Remote Sensing, Vol. 2, Annual Report, Research Bulletin No. 832, Purdue University Agricultural Experimental Station, July 1967, pp. 44-47.

17. Sebestyen, G.S., "Recognition of Membership in Classes", IRE Transactions on Information Theory, Jan. 1961, pp. 44-50.

18. Steinberg, C.A., Abraham, S., and Caceres, C.A., "Pattern Recognition in the Clinical Electrocardiogram", Topics in Computer Techniques for Electrocardiography, Cutler-Hammer, Inc., 1967, pp. 18-25.

19. Wozencraft, J.M., and Jacobs, I.M., Principles of Communication Engineering, John Wiley & Sons, Inc., 1967, pp. 170.

20. Young, T.Y., and Huggins, W.H., "Computer Analysis of Electrocardiograms Using a Linear Regression Technique", IEEE Transactions on Bio-Medical Engineering, July 1964, pp. 60-67.